การแยกประเภทของเสียงดนตรีโดยใช้พาทิเคิลสวอมและ Stacking Ensemble

MUSIC GENRE CLASSIFICATION OF AUDIO SIGNALS USING PARTICLE SWARM OPTIMIZATION AND STACKING ENSEMBLE

กฤติกา เลิศพันธุลักษณ์
KRITTIKA LEARTPUNTULAK

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมอิเล็กทรอนิกส์
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ.2562

KMITL-2019-EN-M-040-124

การแยกประเภทของเสียงดนตรีโดยใช้พาทิเคิลสวอมและ Stacking Ensemble

MUSIC GENRE CLASSIFICATION OF AUDIO SIGNALS USING PARTICLE SWARM OPTIMIZATION AND STACKING ENSEMBLE

กฤติกา เลิศพันธุลักษณ์

KRITTIKA LEARTPUNTULAK

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมอิเล็กทรอนิกส์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ.2562

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การแยกประเภทของเสียงดนตรีโดยใช้พาทิเคิลสวอมและ Stacking Ensemble

MUSIC GENRE CLASSIFICATION OF AUDIO SIGNALS USING PARTICLE SWARM OPTIMIZATION AND STACKING ENSEMBLE

กฤติกา เลิศพันธุลักษณ์

KRITTIKA LEARTPUNTULAK

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมอิเล็กทรอนิกส์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ.2562

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

MUSIC GENRE CLASSIFICATION OF AUDIO SIGNALS USING PARTICLE SWARM OPTIMIZATION AND STACKING ENSEMBLE

KRITTIKA LEARTPUNTULAK

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF ENGINEERING IN ELECTRONICS ENGINEERING

FACULTY OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2019

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2019

FACULTY OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การแยกประเภทของเสียงดนตรีโดยใช้พาทิเคิลสวอม และ Stacking Ensemble
นักศึกษา	นางสาวกฤติกา เลิศพันธุลักษณ์
รหัสประจำตัว	60601148
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมอิเล็กทรอนิกส์
พ.ศ.	2562
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผศ.ดร.ยุทธนา คิดใจเดียว

บทคัดย่อ

การจำแนกประเภทเพลงเป็นกระบวนการของการจัดกลุ่มความคล้ายคลึงกันเช่นรูปแบบสไตล์หรือวัตถุประสงค์ด้วยข้อมูลการจัดการตามที่มีอยู่แล้ว (เช่น ป๊อปและร็อค) ซึ่งถูกใช้พร้อมกับการจำแนกประเภทของหัวข้อ

งานวิจัยนี้จะจัดประเภทเพลงจากสัญญาณเสียงไปยังลำดับชั้นของประเภทดนตรีโดยใช้การแยกลักษณะของ Trimbral texture, Rhythmic content และ Pitch content ถูกใช้เป็นคุณสมบัติหลัก เนื่องจากการเลือกคุณสมบัติที่โดดเด่นถูกเลือกโดยใช้ Particle Swarm Optimization (PSO) จากนั้นนำลักษณะที่ถูกเลือกแล้วไปจำแนกประเภท ผลลัพธ์ความถูกต้องของการจำแนกประเภทไม่ดีนัก ดังนั้นจึงใช้วิธีการ Stacking Ensemble เข้ามาช่วยเพื่อปรับปรุงการทำนายให้ดีขึ้น

ในงานวิจัยนี้มีวัตถุประสงค์เพื่อปรับปรุงการทำนายโดยใช้วิธีการของ Stacking Ensemble ซึ่ง Stacking Ensemble มีสองชั้น คือตัวแยกประเภทพื้นฐาน (Base Classifier) และตัวแยกประเภทเมตา (Meta Classifier) โดยในชั้น Base Classifier ใช้ตัวจำแนกประเภท 5 ชนิดได้แก่ K-Nearest Neighbors (k-NN), ต้นไม้ตัดสินใจ (Decision Tree), Random Forest, Support Vector Machines (SVM) และ Naïve Bayes กระบวนการนี้สร้างขึ้นเพื่อรวมตัวจำแนกหลายๆ ประเภทเข้าด้วยกันแล้วส่งผลลัพธ์จากการทำนายใน Base Classifier ไปยังตัวจำแนกเมตา (Meta Classifier) ในกระบวนการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ของ Meta Classifier จะสร้างโมเดลใหม่เพื่อทำนายข้อมูลการทดสอบ โมเดลใหม่นี้ถูกสร้างขึ้นจากโครงข่ายประสาทเทียม (Neural Network) ซึ่งข้อมูลการฝึกอบรมคือผลลัพธ์จากการทำนายในชั้น Base Classifier

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis	Music Genre Classification based on audio signals Using Particle Swarm Optimization and Stacking Ensemble
Student	Ms. Krittika Leartpantulak
Student ID.	60601148
Degree	Master of Engineering
Program	Electronics Engineering
Year	2019
Thesis Advisor	Asst. Prof. Dr. Yuttana Kitjaidure

ABSTRACT

Genre classification is a process of grouping similarities, such as patterns, styles, or objectives with management data as already in the music (e.g. pop and rock). It is used along with the classification of topics.

This paper will classify songs from the audio signals to a hierarchy of musical genres by using feature extraction. Trimbral texture, rhythmic content and pitch content has been used as the main feature sets. Feature selection was selected by using Particle Swarm Optimization (PSO) and sent selected feature to classification. The result in classification has low accuracy. Thus, using stacking ensemble method is to improve the prediction.

In this paper, the purpose is to improve the prediction by using stacking ensemble method. Stacking ensemble has consists of two levels, base classifier and meta-classifier. In base classifier consists of 5 classification; K-Nearest Neighbors (k-NN), Decision Tree (DT), Random Forest, Support Vector Machines (SVM), and Naïve Bayes. This process was generated to build multiple classifier predictors and sent to meta-classifier. In the

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

process of meta-classifier will create a new model to predict test data. The new model has been created from neural network which train data is the output of base classifier.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จลุล่วงได้ด้วยดีด้วย เนื่องจากได้รับการสนับสนุนและความช่วยเหลือจากอาจารย์ที่ปรึกษาขอขอบพระคุณ ผศ.ดร.ยุทธนา คิดใจเดียว ที่ให้คำแนะนำ คำปรึกษา ตรวจทาน แก้ไขข้อบกพร่องของงานและรายงาน รวมไปถึงช่วยแก้ไขปัญหาต่างๆ ขอขอบพระคุณผู้สนับสนุนหลัก คือบิดา มารดา ที่อำนวยความสะดวกในเรื่องค่าใช้จ่ายและคอยให้กำลังใจ ขอขอบคุณพี่ๆ และเพื่อนๆ ในห้องโปรเจกที่ช่วยให้คำแนะนำ ให้ความรู้และช่วยเหลือ แก้ไข และบุคคลอื่นๆ ที่ไม่ได้กล่าวถึงทุกท่านที่ได้ช่วยเหลือในการทำงานวิจัยนี้

กฤติกา เลิศพันธุลักษณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I
บทคัดย่อภาษาอังกฤษ III
กิตติกรรมประกาศ V
สารบัญ VI
สารบัญตาราง IX
สารบัญรูป X
บทที่ 1 บทนำ 1
1.1 ความเป็นมาและความสำคัญของปัญหา 1
1.2 ความมุ่งหมายและวัตถุประสงค์ 1
1.3 สมมุติฐานของการศึกษา 1
1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย 1
1.5 ขอบเขตของการศึกษา 2
1.6 ขั้นตอนของการศึกษา 2
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง 3
2.1 ประวัติ 3
2.1.1 เพลงลูกทุ่ง 4
2.1.2 เพลงลูกกรุง 6
2.1.3 EDM (Electronic Dance Music) 6
2.1.4 Reggae 7
2.1.5 Music Box 8
2.1.6 เพลงสตริง 8
2.2 งานวิจัยที่เกี่ยวข้อง 9
2.3 การแยกคุณลักษณะ (feature extraction) 9
2.3.1 Rhythmic Content Features (Beat) 10
2.3.2 Timbral Texture Features 10
2.3.2.1 สเปกทรัลเซนทรอยด์ (Spectral Centroid) 10
2.3.2.2 สเปกทรัลโรลออฟ (Spectral Rolloff) 10
2.3.2.3 สเปกทรัลฟลักซ์ (Spectral Flux) 11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

หน้า

2.3.2.4 ซีโร่ครอสซิ่ง (Zero Crossings) 11

2.3.2.5 ความเข้มเสียง 11

2.3.3 ระดับเนื้อเสียง (Pitch Content Features) 11

2.3.4 การแปลงฟูริเยร์ช่วงเวลาสั้น (Short Time Fourier Transform) 12

2.3.5 การแปลงเวฟเลต (Wavelet Transform) 13

2.4 การคัดเลือก feature (feature selection) 14

2.4.1 การหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization) 15

2.5 วิธีการแยกประเภท (Classification) 17

2.5.1 K-nearest neighbor (kNN) 17

2.5.2 Decision Tree classifiers (DT) 18

2.5.3 Random Forest 19

2.5.4 Support Vector Machine (SVM) 20

2.5.5 Naive Bayes 21

2.5.6 Neural Network 21

2.5.6.1 เซลล์ประสาทเทียม 22

2.5.6.2 Model Neural Network 22

2.5.6.3 โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (Back Propagation) 23

2.5.6.4 การเรียนรู้ของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ 23

2.6 Stacking Ensemble 25

บทที่ 3 วิธีดำเนินการวิจัย 27

3.1 ข้อมูลเพลง (Data set) 27

3.2 การดึงคุณลักษณะ (Feature Extraction) 27

3.3 วิธีการคัดเลือกคุณลักษณะ (Feature selection) 32

3.4 Stacking Ensemble 33

3.4.1 Base Classifier 34

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

3.4.2 วิธีการของ k-fold cross validation 34

3.4.3 Meta Classifier 35

3.5 Neural Network 36

บทที่ 4 การทดลองและผลการทดลอง 38

4.1 Feature Extraction 38

4.2 Feature Selection 48

4.3 Stacking Ensemble 55

บทที่ 5 การวิจารณ์หรืออภิปรายผล (DISCUSSION) 61

เอกสารอ้างอิง 62

ภาคผนวก 64

ประวัติผู้เขียน 71

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

3.1 ฟีเจอร์ต่างๆ ที่ถูกนำมาใช้แยกคุณลักษณะของข้อมูล 30

4.1 แสดงค่าความถูกต้องของการใช้ฟีเจอร์ทั้งหมดนำมาแยกประเภทด้วยตัวแยกประเภทชนิดต่างๆ .. 47

4.2 แสดงค่าเปรียบเทียบของตัวจำแนกประเภทต่างๆ ค่า population
และจำนวนการเลือกฟีเจอร์ที่ค่าต่างๆ 50

4.3 แสดงค่าความถูกต้องของ fitness ที่ตัวจำแนกประเภทต่างๆ ที่ค่า population เท่ากับ 20
และจำนวนการเลือกฟีเจอร์เท่ากับ 7 52

4.4 เปรียบเทียบค่าความถูกต้องของตัวจำแนกประเภทต่างๆ ระหว่างใช้ฟีเจอร์ทั้งหมด
และใช้ฟีเจอร์ที่เลือกมาจากวิธีการของ PSO 52

4.5 แสดงค่าความถูกต้องของการสร้างโมเดลจากการทำ 10-fold cross validation 55

4.6 ค่าความถูกต้องของการจำแนกประเภทด้วยวิธีต่างๆ 59

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่ หน้า

2.1 Shot Time Fourier Transform 13

2.2 การวิเคราะห์สัญญาณด้วยการแปลงเวฟเลต 13

2.3 K-nearest neighbor ที่ค่า K ต่างๆ โดยมีจุด X เป็นตัวเทียบ 18

2.4 แสดงตัวอย่างของการตัดสินใจแบบต้นไม้ (Decision Tree) 19

2.5 แสดงตัวอย่างของ Random Forest โดยนำต้นไม้หลายๆ ต้นมาทำการ vote 20

2.6 รูปแบบพื้นฐาน Neural Network 21

2.7 แบบจำลองเซลล์ประสาทเทียม 22

2.8 โครงสร้างของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (Back Propagation) 23

2.9 ขั้นตอนการทำงานของ Stacking Ensemble 26

3.1 แผนภาพการแยกคุณลักษณะของข้อมูล 27

3.2 ด้านซ้ายคือกราฟหน้าต่าง Hamming Window และ ด้านขวาคือ Fourier Transform ของ Hamming Window 28

3.3 Mel Scale Filter Bank 28

3.4 การแปลงเวฟเลตแบบเต็มหน่วย (Discrete wavelet transform) 29

3.5 แผนภาพแสดงการคำนวณหา Beat Histogram 29

3.6 ขั้นตอนกระบวนการทำ PSO 32

3.7 แผนภาพแสดงวิธีการของ Stacking Ensemble 33

3.8 วิธีการทำงานของ Base Classifier 34

3.9 วิธีการทำงานของ k-fold cross validation 35

3.10 วิธีการทำงานของ Meta Classifier 35

3.11 วิธีการทำงานของ Neural Network 36

3.12 โครงข่ายของ Neural Network 37

4.1 ค่าที่ได้จากฟีเจอร์ Mean Spectral Centroid 39

4.2 ค่าที่ได้จากฟีเจอร์ Mean Spectral Flux 39

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่ หน้า

4.3 ค่าที่ได้จากฟีเจอร์ Mean Spectral Rolloff 39
4.4 ค่าที่ได้จากฟีเจอร์ Mean Volume 40
4.5 ค่าที่ได้จากฟีเจอร์ Mean Zero Crossing 40
4.6 ค่าที่ได้จากฟีเจอร์ Mean Short Time Energy 40
4.7 ค่าที่ได้จากฟีเจอร์ Low Energy 41
4.8 ค่าที่ได้จากฟีเจอร์ Std First MFCC 41
4.9 ค่าที่ได้จากฟีเจอร์ Std Second MFCC 41
4.10 ค่าที่ได้จากฟีเจอร์ Std Third MFCC 42
4.11 ค่าที่ได้จากฟีเจอร์ Std Fourth MFCC 42
4.12 ค่าที่ได้จากฟีเจอร์ Std Fifth MFCC 42
4.13 ค่าที่ได้จากฟีเจอร์ Period Peak Unfolded Pitch 43
4.14 ค่าที่ได้จากฟีเจอร์ Period Peak Folded Pitch 43
4.15 ค่าที่ได้จากฟีเจอร์ Strength Pitch 43
4.16 ค่าที่ได้จากฟีเจอร์ BPM 44
4.17 ค่าที่ได้จากฟีเจอร์ Ratio Amplitude Histogram Beat 44
4.18 ค่าที่ได้จากฟีเจอร์ Beat Strength 44
4.19 ค่าที่ได้จากฟีเจอร์ Mean Spectral Contrast 45
4.20 ค่าที่ได้จากฟีเจอร์ Mean Spectral Mean 45
4.21 ค่าที่ได้จากฟีเจอร์ Mean of Frequency Histogram 45
4.22 ค่าที่ได้จากฟีเจอร์ Mean of Pitch 46
4.23 ค่าที่ได้จากฟีเจอร์ Std of Pitch 46
4.24 ค่าที่ได้จากฟีเจอร์ Mean of Pitch Strength 46
4.25 ค่าที่ได้จากฟีเจอร์ Std of Pitch Strength 47
4.26 ค่าที่ได้จากฟีเจอร์ Sum of Pitch Strength 47

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่ หน้า

4.27 ค่าความถูกต้องของสมการ fitness โดยใช้ K-Nearest Neighbor 48
4.28 ค่าความถูกต้องของสมการ fitness โดยใช้ Support Vector Machine 49
4.29 ค่าความถูกต้องของสมการ fitness โดยใช้ Decision Tree 49
4.30 ค่าความถูกต้องของแต่ละ classifier ที่จำนวนการเลือกฟีเจอร์ที่ต่างกัน
โดยมีค่า population ที่ 20 51
4.31 ค่าความถูกต้องของแต่ละ classifier ที่จำนวนการเลือกฟีเจอร์เท่ากับ 7 และ
population เท่ากับ 20 51
4.32 Confusion Matrix ของ Decision Tree ที่ผ่านการเลือกใช้ฟีเจอร์ (PSO) 52
4.33 Confusion Matrix ของ K-Nearest Neighbor ที่ผ่านการเลือกใช้ฟีเจอร์ (PSO) 53
4.34 Confusion Matrix ของ Support Vector Machine ที่ผ่านการเลือกใช้ฟีเจอร์ (PSO) 53
4.35 Confusion Matrix ของ Random Forest ที่ผ่านการเลือกใช้ฟีเจอร์ (PSO) 53
4.36 Confusion Matrix ของ Decision Tree ที่ใช้ฟีเจอร์ทั้งหมด 54
4.37 Confusion Matrix ของ K-Nearest Neighbor ที่ใช้ฟีเจอร์ทั้งหมด 54
4.38 Confusion Matrix ของ Support Vector Machine ที่ใช้ฟีเจอร์ทั้งหมด 54
4.39 Confusion Matrix ของ Random Forest ที่ใช้ฟีเจอร์ทั้งหมด 55
4.40 Confusion Matrix ของ Decision Tree 56
4.41 Confusion Matrix ของ K-Nearest Neighbor 56
4.42 Confusion Matrix ของ Support Vector Machine 56
4.43 Confusion Matrix ของ Random Forest 57
4.44 Confusion Matrix ของ Naïve Bay 57
4.45 Confusion Matrix ของตัวจำแนกประเภท Neural Network แบบ Pattern Recognition 58
4.46 Confusion Matrix ของการทำ Stacking Ensemble 59

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การประเภทของเพลงโดยใช้วิธีการฟังเป็นวิธีการดั้งเดิมในการวิเคราะห์ข้อมูลดนตรี ซึ่งวิธีการดังกล่าวไม่สามารถทำได้ในกรณีที่มีเพลงเป็นจำนวนมาก อีกทั้งประสบการณ์ของผู้ฟังส่งผลต่อความถูกต้องของการวิเคราะห์ข้อมูลโดยตรง เนื่องด้วยข้อจำกัดที่กล่าวมาแล้ว วิธีการค้นหาข้อมูลเพลงด้วยเทคนิคทางคอมพิวเตอร์จึงถูกนำมาประยุกต์ใช้ในการวิเคราะห์ข้อมูลดนตรีแทนการฟัง เทคนิคทางการรู้จำดนตรีสามารถช่วยจัดกลุ่ม แยกกลุ่ม และระบุลักษณะเฉพาะของงานดนตรีได้ โดยกระบวนการรู้จำดนตรีจะดึงข้อมูลจากเพลงเพื่อเป็นข้อมูลเข้าและหารูปแบบ (Pattern) ที่เหมือนหรือคล้ายกันของข้อมูลกลุ่มนั้นๆ เพื่อนำมาใช้ในการวิเคราะห์ ตัวอย่างของการนำความรู้เรื่องการรู้จำดนตรีมาประยุกต์ใช้ เช่น การตรวจหาการละเมิดลิขสิทธิ์ การระบุข้อมูลเพลงในกรณีที่ไม่ทราบแหล่งที่มา การระบุประเภทเครื่องดนตรีจากเสียง และการแยกประเภท เป็นต้น

1.2 ความมุ่งหมายและวัตถุประสงค์

เพื่อจัดแยกประเภทแนวเพลง หรือหมวดหมู่ของดนตรีที่มีปริมาณมาก เพื่อให้ง่ายต่อการค้นหาและนำไปใช้งานได้ง่าย

1.3 สมมุติฐานของการศึกษา

นำกระบวนการของ Machine learning มา ใช้ปรับปรุงการทำนายผลลัพธ์ของการแยกประเภทเพลง โดยจะใช้วิธีการคัดเลือกคุณลักษณะที่เด่นๆ ของฟีเจอร์ และ Stacking Ensemble มาช่วยปรับปรุงคุณภาพการทำนายให้ดียิ่งขึ้น

1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย

ทฤษฎีที่จำเป็นในงานวิจัยนี้ ได้แก่ การแยกคุณลักษณะ (Feature Extraction), การคัดเลือกคุณลักษณะที่เด่นๆ ออกมา (Feature Selection) และการปรับปรุงคุณภาพการทำนายของการแยกประเภทแนวเพลง โดยใช้วิธีการ Ensemble เข้ามาช่วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.5 ขอบเขตของการศึกษา

งานวิจัยนี้ได้พัฒนาต่อยอดมาจากงานวิจัยก่อนๆ เพื่อปรับปรุงการทำนายให้ดีมากขึ้น โดยใช้วิธีการ Ensemble เข้ามาช่วยปรับปรุงการทำนาย โดยนำเอากระบวนการของ Machine learning การเรียนรู้แบบต่างๆ เข้ามาผสมผสานกัน โดยใช้มากกว่า 1 การเรียนรู้ เพื่อทำให้การแยกประเภทของเพลงมีประสิทธิภาพมากขึ้น

1.6 ขั้นตอนของการศึกษา

ศึกษาทฤษฎีและงานวิจัยต่างๆ ที่เกี่ยวข้องกับการแยกประเภทของเสียง เช่น การแยกฟีเจอร์, การเลือกฟีเจอร์ที่เด่นๆมาใช้งาน และวิธีการปรับปรุงการทำนายผลของข้อมูลในบทที่ 2 จากนั้นนำหลักการทั้งหมดมาใช้กับข้อมูลเพลงที่ได้เตรียมไว้ตามขั้นตอน ในบทที่ 3 แล้วทำการทดสอบการทำงานของระบบที่ได้ออกแบบ ในบทที่ 4 ทำการสรุปผลทั้งหมดในบทที่ 5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

2.1 ประวัติ

ดนตรีเป็นงานศิลป์อีกแขนงหนึ่งที่มนุษย์ได้รับอิทธิพลจากธรรมชาติ สิ่งแวดล้อมอันกว้างใหญ่เป็นแรงบันดาลใจให้เกิดความคิดสร้างสรรค์ เรียนรู้ เลียนแบบจากธรรมชาติเป็นการตอบสนองความต้องการโดยตรง โดยใช้วัสดุจากธรรมชาติประดิษฐ์เป็นเครื่องดนตรีใช้ในการบรรเลงขับกล่อมตามกิริยาท่าทางของมนุษย์ที่พึงกระทำได้ ได้แก่ ดีด-สี-ตี-เป่า

ในปัจจุบันดนตรียังคงเป็นวัฒนธรรมที่มีบทบาทต่อผู้คนในสังคมไทยหลายส่วน ตั้งแต่ในชีวิตประจำวันไปจนถึงเทศกาลประเพณีสำคัญระดับประเทศ ตลอดไปจนถึงสถานบันเทิงที่เปิดให้บริการกันอย่างมากมายยามค่ำคืนในสังคมเมือง อย่างไรก็ตาม การที่เราจะสามารถเข้าใจดนตรีได้จึงไม่ใช่เพียงแต่รู้ว่าในสังคมไทย มีเครื่องดนตรีอะไรบ้าง มีกี่อย่าง กี่ประเภท แต่ต้องอยู่ที่ความเข้าใจเกี่ยวกับสภาพแวดล้อมทางสังคมที่ดนตรีเหล่านั้นแสดงบทบาทของพวกมันอยู่ด้วยและการเข้าใจภูมิหลังทางประวัติศาสตร์ของสังคมนั้นๆ ก็เป็นสิ่งหนึ่งที่จะทำให้เรารู้จักดนตรีเครื่องดนตรี และโอกาสในการเล่นเครื่องดนตรีชิ้นนั้นๆ ได้อย่างลึกซึ้งอีกด้วย

นอกจากนั้นแล้ว ดนตรียังมีบทบาทหน้าที่เป็นสื่อ ที่สะท้อนเรื่องราวต่างๆ ของมนุษย์และโลกมนุษย์เข้าใจเรื่องราวต่างๆ ในโลก เรียนรู้ ชีวิต เรื่องราวของผู้คนและความสัมพันธ์ของปรากฏการณ์ต่างๆ จากการถ่ายทอดผ่านบทเพลง ซึ่งสามารถใช้ถ่ายทอดเรื่องราวเหล่านี้ได้เป็นอย่างดี ซึ่งการสื่อสารในที่นี้หมายถึง กระบวนการที่ศิลปินได้แสดงออกถึงอารมณ์ ความรู้สึก ความนึกคิด และจินตนาการของตนเองออกมาอย่างเป็นรูปธรรม เพลงจึงเป็นผลผลิตของแรงผลักดันอันเกิดจากแรงบันดาลใจ ที่หล่อหลอมขึ้นจากประสบการณ์ชีวิตและความเป็นมนุษย์นั่นเอง

เพลงนั้นน่าจะเกิดขึ้นมาพร้อมๆกับมนุษย์ โดยอาจจะเกิดจากตัวมนุษย์เอง ทุกครั้งที่มนุษย์มีการเปล่งเสียงออกมาจะเกิดทำนองอะไรก็ได้ หรืออาจจะเกิดจากสภาพแวดล้อม ความเป็นอยู่ ตลอดทั้งการดำรงชีวิตของมนุษย์ สิ่งเหล่านี้น่าจะเป็นแรงบันดาลใจให้เกิดเพลง โดยเสียงที่เปล่งออกมานั้นอาจมีความหมายหรือไม่มีความหมายก็ได้ เมื่อนำเสียงมนุษย์ผสมผสานกับจังหวะท่วงทำนองที่เลียนเสียงธรรมชาติ ทำให้เกิดเสียงเพลงหรือเสียงดนตรีที่มีความไพเราะงดงาม และมีการพัฒนาขึ้นตามลำดับ

น.พ.พูนพิศ อมาตยกุล ได้ตั้งข้อสันนิษฐานเกี่ยวกับการเกิดของเพลงไทยไว้ดังนี้ “ดนตรีนั้นเป็นภาษาชนิดหนึ่งของมนุษย์เรานี้ ชาติใดก็ตามที่มีภาษาพูด ภาษาเขียนของตนเองก็มักจะต้องมีดนตรีเป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ของตนเองด้วย เพราะดนตรีและเพลงร้องนั้นเป็นส่วนหนึ่งที่พัฒนามาพร้อมๆกับภาษาพูด " (พูนพิศ อมาตยกุล 2519:1)

เพลงไทย (Thai Traditional Music)หรือเพลงไทยเดิมนั้น ยังไม่มีข้อสรุปว่าเกิดขึ้นและสืบทอดกันมาแต่เมื่อไร อย่างไรก็ดีท่านอาจารย์มนตรี ตราโมท ได้ตั้งข้อสันนิษฐานว่าเพลงไทยนั้นได้เริ่มต้นมาตั้งแต่สมัยสุโขทัยเป็นต้นมา (มนตรี ตราโมท 2515 : 15)

ส่วนดนตรีตะวันตกได้เริ่มเข้าสู่ประเทศไทยอย่างจริงจังในรัชสมัยพระบาทสมเด็จพระจอมเกล้าเจ้าอยู่หัว เมื่อราวพ.ศ.2394-2411 โดยที่ได้โปรดเกล้าฯให้มีการจัดตั้งกองทหารฝึกแบบยุโรปขึ้น และการเดินแถวของทหารฝรั่งนั้นมีแตรเป่านำหน้า คนไทยจึงเริ่มคุ้นเคยกับแตรฝรั่งมาแต่ครั้งนั้น (พูนพิศ อมาตยกุล 2529:14)

ตั้งแต่บัดนั้นมาอารยะธรรมตะวันตกเข้ามาผูกพันกับวัฒนธรรมดั้งเดิมของไทย เมื่อเพลงไทยได้รับอิทธิพลจากดนตรีฝรั่ง เพลงไทยจึงได้ขยายตัวทำให้เกิดรูปแบบใหม่ขึ้น เรียกว่า "เพลงไทยสากล" ซึ่งเพลงไทยสากลในยุคต้นๆ กองทัพเรือและกองทัพบกได้ริเริ่มนำเพลงไทยเดิมมาบรรเลงด้วยเครื่องเป่าลมไม้และเครื่องเป่าทองเหลือง ซึ่งต่อมามีชื่อเรียกวงประเภทนี้ว่า "โยธวาฑิต"(สุมณฑา ศักดิ์ชัยสมบูรณ์ 2540 : 17)

วิวัฒนาการทางดนตรีไทยและเพลงไทยมีพัฒนามาอย่างต่อเนื่อง เริ่มจากการประดิษฐ์เครื่องดนตรีไทย การประสมวงดนตรีไทยและการขับร้องเพลงไทย ซึ่งแสดงถึงศิลปะทางดนตรีที่เป็นเอกลักษณ์ของไทยที่เรียกว่า ดนตรีไทยและเพลงไทย

เพลงไทย หมายถึงเพลงที่แต่งขึ้นตามหลักของดนตรีไทย มีลีลาในการขับร้องและบรรเลงแบบไทยโดยเฉพาะและแตกต่างจากเพลงของชาติอื่นๆ เพลงไทยแต่เดิม มักจะมีประโยคสั้นๆและมีจังหวะค่อนข้างเร็ว ส่วนใหญ่มีต้นกำเนิดมาจากเพลงพื้นบ้าน หรือเพลงสำหรับประกอบการรำเต้นเพื่อความสนุกสนานรื่นเริง ต่อมาเมื่อต้องการจะใช้เป็นเพลงสำหรับร้องขับกล่อม และประกอบการแสดงละครก็จำเป็นต้องประดิษฐ์ทำนองให้มีจังหวะช้ากว่าเดิม และมีประโยคยาวกว่าเดิม ให้เหมาะสมที่จะร้อง

2.1.1 เพลงลูกทุ่ง

เพลงที่สะท้อนวิถีชีวิต สภาพสังคม อุดมคติและวัฒนธรรมไทย มีท่วงทำนอง คำร้อง สำเนียง ลีลาการร้อง การบรรเลง ที่เป็นแบบแผน มีลักษณะเฉพาะ ซึ่งให้บรรยากาศของความเป็นลูกทุ่ง โดยเฉพาะการร้องเอื้อนที่ใช้ลูกคอ

เพลงลูกทุ่งเป็นเพลงไทยที่มีลักษณะเฉพาะ กล่าวคือ มีการใช้ภาษาง่ายๆ บรรยายเรื่องราวของชีวิตสภาพสังคมและวัฒนธรรม ซึ่งเน้นชีวิตของชาวชนบท โดยก่อนหน้านั้นยังไม่มีการแบ่งแยกว่า เป็นลูกทุ่งหรือลูกกรุง เนื่องจาก เพลงไทยได้พัฒนามาจากเพลงไทยเดิมที่ใช้คำร้องที่มีลักษณะการเอื้อน และใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เครื่องดนตรีไทยบรรเลง มาเป็นเพลงไทยสากล ที่มีเนื้อร้องเต็ม ใช้เครื่องดนตรีสากลบรรเลง ซึ่งได้รับความนิยมมากขึ้น เนื่องจากฟังง่าย ร้องง่าย จำได้ง่าย และมักเรียกลักษณะจากแนวการร้องของนักร้องแต่ละคนมากกว่า เช่น แนวรำวง แนวเพลงเพื่อชีวิต นอกจากนี้ นักประพันธ์เพลงรุ่นเก่าๆ ก็ไม่เห็นด้วยที่จะให้มีการแบ่งแยกประเภทของเพลงไทยสากลว่า เป็นเพลงลูกทุ่ง และลูกกรุง เพราะถือว่า ดนตรีเป็นภาษาสากลและต่างเป็นเพลงไทยสากลเช่นเดียวกัน

คำว่า "ลูกทุ่ง" ปรากฏครั้งแรกเป็นชื่อของภาพยนตร์เพลง สร้างโดยบริษัทภาพยนตร์ไทย หรือไทยฟิล์ม เมื่อ พ.ศ. 2481 โดยมี พระเจ้าวรวงศ์เธอ พระองค์เจ้าภาณุพันธุ์ยุคล เป็นผู้กำกับ ผู้แต่งทำนองเพลง คือ หม่อมหลวง พวงร้อย อภัยวงศ์ ผู้แต่งเนื้อร้องคือ พระยาโกมารกุลมนตรี (ชื่น) เพลงในภาพยนตร์ได้แก่ เงาไม้ ต้อนกระบือ เกี้ยวสาว สายัณห์ และไม้งาม เนื้อร้องส่วนใหญ่เน้นการชมธรรมชาติอันสวยงามของท้องไร่ปลายนา

เพลงที่มีแนวเป็นเพลงลูกทุ่ง และบันทึกไว้เป็นหลักฐานคือ เพลง สาวชาวไร่ ซึ่งเป็นผลงานการประพันธ์ของ ครูเหม เวชกร ใช้ร้องประกอบการแสดงละครวิทยุเรื่อง สาวชาวไร่ ใน พ.ศ. 2481 จากหลักฐานที่ค้นพบมีเค้าที่น่าเชื่อว่า เพลงสาวชาวไร่ เป็นเพลงลูกทุ่งเพลงแรกของเมืองไทย

การตีความเรื่องกำเนิดของเพลงลูกทุ่งยังมีความเห็นแตกต่างกันอยู่บ้าง กาญจนาคพันธ์ เขียนไว้ในหนังสือ เรื่องของละครและเพลง ว่า "หลังสงครามโลกครั้งที่ 2 แล้ว มหรสพต่างๆ ก็เริ่มฟื้นตัวขึ้น เกิดเพลงแบบใหม่เรียกกันว่า เพลงลูกทุ่ง ตั้งเป็นวงดนตรีแบบสากล" ส่วน พยงค์ มุกดา เห็นว่า เรื่องแผลเก่า ของไม้เมืองเดิม มีเพลงขวัญของเรียม ซึ่งพรานบูรพ์แต่งขึ้นใน พ.ศ. 2482 น่าจะถือเป็นแม่บทของเพลงลูกทุ่งได้ เพราะมีลูกเอื้อน ลูกขัด และสาระเนื้อหาแบบเพลงลูกทุ่ง

ราว พ.ศ. 2483 ป.วรานนท์ นักจัดรายการวิทยุ ได้เรียกนักร้องประเภทเพลงชีวิต เช่น คำรณ สัมบุณณานนท์ สมยศ ทัศนพันธ์ พยงค์ มุกดา สุรพล สมบัติเจริญ ว่าเป็น "นักร้องตลาด" เป็นเพลงตลาดเนื่องจากเข้าถึงผู้ฟังที่เป็นชาวบ้านทั่วไปได้ง่าย เนื้อหาเพลงมีความชัดเจน ส่วนอีกแนวหนึ่ง เรียกว่า "เพลงผู้ดี" ซึ่งหมายถึง เพลงที่มีการแต่งอย่างไพเราะเพราะพริ้ง นักร้องแนวนี้ เช่น สุเทพ วงศ์กำแหง สวลี ผกาพันธุ์ เพลงไทยสากลจึงถูกแบ่งประเภทออกเป็น เพลงลูกทุ่ง และเพลงลูกกรุง แต่จะใช้หลักเกณฑ์ใดในการตัดสินนั้น มีผู้ให้ความเห็นว่า เป็นเพราะช่องว่างระหว่างวัฒนธรรมเมือง และชนบท เพลงลูกกรุงจึงได้รับความนิยมในหมู่ผู้มีการศึกษาของสังคมเมือง ส่วนเพลงลูกทุ่งมุ่งรับใช้ชาวชนบทเป็นส่วนมาก

คำว่า "เพลงลูกทุ่ง" เกิดขึ้นโดย นายจำนง รังสิกุล เป็นผู้ริเริ่ม ในระหว่างที่ดำรงตำแหน่ง เป็นหัวหน้าฝ่ายจัดรายการ ของสถานีโทรทัศน์ช่อง 4 บางขุนพรหม โดยตั้งชื่อรายการว่า "เพลงลูกทุ่ง" มีนายประกอบ ไชยพิพัฒน์ เป็นผู้จัด ออกอากาศ เมื่อ พ.ศ. 2507 เดือนละ 2 ครั้ง ทุกวันจันทร์เว้นจันทร์ จน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ได้รับความนิยมและแทรกซึมไปตามวงดนตรีต่างๆ เริ่มเรียกวงของตนเองว่า วงดนตรีลูกทุ่ง และนักร้องนิยมเรียกตัวเองว่า นักร้องเพลงลูกทุ่ง

2.1.2 เพลงลูกกรุง

เพลงลูกกรุงกำเนิดขึ้นเมื่อประมาณ พ.ศ. 2474 ซึ่งเป็นช่วงที่ประเทศไทยกำลังจะเปลี่ยนแปลงระบอบการปกครองฯ ในรัชกาลที่ 7 บางแหล่งบอกปลาย รัชกาลที่ 6 ประมาณ พ.ศ. 2455 โดยการจ้างครูจากอิตาลีนำเครื่องสายสากลเข้ามาสอน แต่เพลงลูกกรุงเริ่มชัดเจนเมื่อประมาณ ปี พ.ศ. 2475 โดยเริ่มมีแนวเพลง เนื้อร้อง ทำนอง และเครื่องดนตรีที่นำมาบรรเลงประกอบเป็นไทย นำทำนองเพลงของรัชกาลที่ 6 ที่นำมาเรียบเรียงเสียงประสานใช้เครื่องดนตรีสากลบรรเลงประกอบ เป็นเพลงในกิจการลูกเสือต่อมาได้มีการนำเพลง "ลาทีกล้วยไม้" ของขุนวิจิตรมาตรา มาทำในจังหวะรุมบ้าเพลงแรกของไทย และบทเพลงของจิตร ภูมิศักดิ์ และอื่นๆ ซึ่งแต่ง และใช้เครื่องดนตรีสากลบรรเลงประกอบ เป็นเครื่องบ่งบอกการกำเนิดของเพลงลูกกรุง

เพลงลูกกรุง เป็นเพลงไทยสากลประเภทหนึ่ง โดยเป็นเพลงที่บอกเล่า ถ่ายทอด ความรู้สึกของสังคม และคนเมืองหลวง ตลอดจนเหตุการณ์ที่เกิดขึ้น การถ่ายทอดอารมณ์ การขับร้อง น้ำเสียง ของกลุ่มนักร้อง นักแต่งเพลง และนักดนตรีจะมีรูปแบบ ประณีต ละเอียดอ่อน ออกมานุ่มนวล เนื้อร้องจะมีลักษณะเป็นร้อยแก้ว ร้อยกรอง มีความหมายสลับซับซ้อน ยอกย้อน ผู้ประพันธ์ นักร้อง นักดนตรี และผู้ฟังอาจเข้าถึงบทเพลงไปตามจินตนาการแตกต่างกันไป

คำว่า “เพลงลูกกรุง” ได้แยกกลุ่มผู้ฟังอย่างเด่นชัดขึ้นมาตามลำดับ โดยนำเอาความเป็นอยู่ (Lifestyle) ของผู้ฟัง ศิลปิน นักร้อง นักดนตรี เป็นผู้กำหนดทิศทาง วงสุนทราภรณ์เป็นวงแรก ก่อตั้งเป็นวงดนตรีวงใหญ่ เมื่อ พ.ศ. 2482 ซึ่งทำให้สังคมเมืองในยุคนั้น เริ่มตื่นตัวการฟังเพลง ผู้ฟังและค่ายเพลงต่างๆ เริ่มจัดประเภทเพลง สร้างนักร้องให้มีรูปแบบความเป็นคนเมืองหลวง นักดนตรี นักแต่งเพลง นักร้องต่างมีรูปแบบ หาแฟชั่นนำสังคม ทั้งเรื่องการแต่งตัวการออกกินข้าวนอกบ้าน มีคลับมีบาร์ แถวถนนราชดำเนิน ตามย่านชุมชน โรงแรมใหญ่ๆ มีห้องบอลรูม เพื่อให้มีการจัดแสดงดนตรีประกอบ กลุ่มคนฟังจึงเริ่มรับแนวเพลงลูกกรุง เกิดการเปรียบเทียบระหว่างแนวเพลงลูกทุ่ง กับเพลงลูกกรุงขึ้นมาโดยอัตโนมัติ

2.1.3 EDM (Electronic Dance Music)

EDM หรือชื่อเต็ม ๆ ว่า Electronic Dance Music เป็นแนวของดนตรีที่ได้รับความนิยมในต่างประเทศเป็นอย่างมาก ขนาดที่ว่ากลายเป็นดนตรีกระแสหลักไปแล้ว ซึ่งแนวดนตรีแบบ EDM ก็จะใช้เครื่องดนตรีอิเล็กทรอนิกส์และเทคโนโลยีเป็นหลัก เล่นซาวด์แบบที่ให้ความรู้สึกสนุก ครื้นเครง อยากจะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลุกขึ้นมาเต้น โดยจะอยู่ในรูปแบบของ Digital ซึ่งทางฝั่งยุโรปฮิตกันมาก สำหรับในเมืองไทยก็จะเป็นแนวดนตรีที่วัยรุ่นเรียกกันติดปากว่า “เพลงตื๊ด” นั่นเอง

ดนตรีแนวอิเล็กทรอนิกส์นี้เกิดขึ้นมาได้นานแล้ว เพียงแต่ว่าพอคนเรียกชื่อกันรวมๆ ก็เลยเรียกเหมาไปหมดว่าเป็น EDM ได้ไม่นานมานี้ โดยคำว่า EDM เป็นชื่อที่พึ่งเลือกใช้ขึ้นมาในช่วง 10 กว่าปีที่ผ่านมานี่เอง จุดเริ่มต้นของเพลงนี้มาจาก House Music ที่เกิดขึ้นทางแถบยุโรปเป็นเพลงที่มาจากวัยรุ่นผิวขาวได้คิดค้นขึ้นมาเนื่องจากว่าไม่ต้องการเพลงแนวเดียวกันกับคนผิวสี ในช่วงปลายยุค 70 ถึงช่วงต้นยุค 80 เป็นช่วงที่คนผิวสีเริ่มได้รับการยอมรับมากขึ้นเรื่อยๆ ในช่วงเวลาดังกล่าวเพลงแนว Soul, Funk, Disco จะถือว่ามาแรงมาก อย่างไรก็ตามในช่วงเวลาดังกล่าวยังมีเรื่องของการเหยียดผิวและต่อต้านคนสีผิวกันอยู่ เพลงแนวนี้จึงเกิดขึ้นมาจากงานปาร์ตี้ในโกดังที่ได้รับความนิยมเป็นอย่างมากจนคนได้มีการเรียกดนตรีแนวอิเล็กทรอนิกส์เหล่านี้ในยุคแรกๆ ว่า House Music ซึ่งเอาเข้าจริงก็ไม่ได้มีชื่อเรียกอย่างเป็นทางการด้วยซ้ำแต่ด้วยความที่มันมาจากการเล่นในโกดังหรือในบ้านนั่นเอง จนบ้านหรือโกดังโดนเผาจากพวกหัวรุนแรงทำให้กระแสเหล่านี้แรงขึ้นเรื่อยๆ จนเข้าไปในฝั่งอเมริกา กลายเป็นการแตกแขนงทางด้านดนตรีแนวต่างๆ อีกมากมาย อาทิ Electro House, Drum & Bass, Trance เป็นต้น จนค่อยๆ ก่อไล่เรียงตามยุคสมัยมาเป็นดนตรีแนวต่างๆ ในยุคปัจจุบันนี้เอง

เพลงแนวต่างๆ เหล่านี้จึงถูกเรียกแบบรวมๆ กันว่า EDM เหมือนอย่างที่เราคุ้นเคยกันดีเป็นเพลงที่มักเปิดในงานปาร์ตี้หรืองานสังสรรค์ต่างๆ เพื่อให้เกิดความสนุกสนานมากยิ่งขึ้นกว่าเดิม เมื่อคนได้ฟังก็จะมีความรู้สึกอยากลุกขึ้นมาเต้น ขึ้นมาโยกย้ายตามจังหวะทำให้งานปาร์ตี้มีสีสัน มีความสนุกสนานยิ่งกว่าเดิม ทุกคนต่างก็มีความสุขเสมอเมื่อได้ฟังเพลงแนวนี้

2.1.4 Reggae

เพลงแนวเร็กเก้นั้นเกิดขึ้นในปี ค.ศ.1968 ในผลงาน Do the Reggay ของกลุ่มศิลปินแนวสกา ทูทส์ แอนด์ เดอะเมย์ทัลส์ (Toots and The Maytals) วงดนตรีแนวเร็กเก้จะมีการใช้เครื่องดนตรีที่ทันสมัยกว่าวงดนตรีแนวสกา และ ร็อคสเตดี้ เช่น กีต้าร์ไฟฟ้า เบสไฟฟ้า เปียโน ออร์แกน กลองชุด เครื่องสังเคราะห์เสียงและมีการนำเอากลองเพอร์คัสชั่น เข้ามาประกอบดนตรีด้วย และความสำคัญของเครื่องเป่าก็ลดน้อยลงเมื่อเทียบกับดนตรีแนวสกาแต่ก็ยังมีใช้อยู่ในบางโอกาส ในปัจจุบันดนตรีสกา-เร็กเก้ได้แพร่กระจายไปยังประเทศต่างๆ โดยกลุ่มเชื้อชาติอื่นๆได้นำเอาสไตล์ดนตรีของสกา-เร็กเก้ ไปใช้ในการสื่อสารประเด็นต่างๆ ที่สอดรับกับบริบทในแต่ละท้องถิ่น และใส่ความหมายใหม่เข้าไปในรูปแบบเฉพาะของตน วัฒนธรรมดนตรีสกา-เร็กเก้เริ่มปรากฏขึ้นในสังคมไทยช่วง พ.ศ.2527-2534 โดยมีศิลปินเพลงเพื่อชีวิตอย่างวงคาราบาว ในชื่อเพลงสองเฒ่าผู้ยิ่งใหญ่ จากอัลบั้มเมดอินไทยแลนด์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นอกจากนี้ยังมี ปานศักดิ์ รังสิพราหมณสกุล ในชื่อเพลงไปทะเลกันดีกว่า จากอัลบั้มไปทะเล เป็นที่แน่ชัดว่าในช่วงนี้แนวดนตรีสกา-เร็กเก้ ยังไม่ค่อยเป็นที่รู้จักและนิยมเท่าไรนักต่อมาในช่วง พ.ศ.2535 ดนตรีสกา-เร็กเก้ ก็ได้ปรากฏในอุตสาหกรรมดนตรีไทยค่อนข้างชัดเจน จากการเริ่มต้นของสองพี่น้อง นครินทร์และเจษฎา ธีระภินันท์ ที่รวบรวมเพื่อนพ้องก่อตั้งวงดนตรี โดยการเป็นวงเล่นประจำตามร้านต่างๆ จึงเป็นที่มาในการเกิดวงทีโบน (T-Bone)

หลังจากนั้นจึงทำอัลบั้มของตัวเอง และนั่นเป็นการเปิดตัวสกา-เร็กเก้ให้ชาวไทยได้ทำความรู้ จักอย่างใกล้ชิด ในนามวงดนตรีแนวสกา-เร็กเก้ กลุ่มแรกของประเทศไทย ในอัลบั้มชุดจังหวะนี้ใจดีเข้ากระดูกดำ ในสังกัดค่ายเพลง มูเซอร์ เร็คคอร์ด ซึ่งถือเป็นการบุกเบิกถนนดนตรีแนวสกา-เร็กเก้ในเมืองไทยอย่างเป็นทางการ ด้วยเนื้อหาดีๆ ในบทเพลง พร้อมความแปลกใหม่ของแนวดนตรี ทำให้วงการเพลงเมืองไทยเริ่มมีสีสันมากขึ้น จากที่มีแต่กระแสดนตรีของค่ายใหญ่เพียงไม่กี่ค่าย โดยมีเพลงฮิตอย่างเธอเห็นท้องฟ้านั่นไหม เพลงที่สร้างชื่อให้กับวงทีโบนจนเป็นที่รู้จักในวงการเพลงเมืองไทยในยุคนั้น

2.1.5 Music Box

กล่องดนตรี หรือ Music box มีต้นกำเนิดมาจากระฆังในโบสถ์ที่ใช้ตีเพื่อบอกเวลา คล้ายๆ กับระฆังวัดบ้านเรา แต่ระฆังโบสท์แต่ละใบจะมีเสียงสูงต่ำแตกต่างกัน เวลาตีแล้วจะได้ยินเป็นเสียงเพลงเพลงหนึ่งที่ดังมากๆ และยังถูกนำมาใช้กันอย่างแพร่หลายมาจนทุกวันนี้ก็คือ เพลงเวสมินเตอร์ ชาร์ม (Westminster Chimes) นั่นเอง

ปี ค.ศ. 1796 นาย อองตวน ฟาเวร่ (Antoine Favre) ชาวเมืองเจนีวา ประเทศสวิตเซอร์แลนด์นำเทคนิคสร้างเสียงเพลงจากระฆังนี้มาดัดแปลง โดยใช้แท่งโลหะและลูกตุ้มไปติดแทนแล้วเชื่อมโยงด้วยหมุดเหล็ก พัฒนาให้เป็นนาฬิกาเสียงดนตรี ปี ค.ศ. 1802 นายอองตวน ฟาเวร่ ได้นำผลงานประดิษฐ์ดังกล่าวมาย่อส่วนใส่ลงในกล่องยานัตถุ์ เป็นกล่องยานัตถุ์เพลง (Music snuff box) ซึ่งถือเป็นต้นแบบของกล่องดนตรีในเวลาต่อมา

ในช่วงปี ค.ศ. 1850-1870 ถือเป็นช่วงที่มีการประดิษฐ์กล่องดนตรีที่ประณีตที่สุดทั้งในด้านเสียงเพลงและตัวกล่อง จากนั้นในช่วงปี ค.ศ. 1880 อุตสาหกรรมการผลิตกระบอกโลหะ (Cylinder) เฟื่องฟูมาก เมื่อกล่องเพลงเป็นที่นิยมแพร่หลาย ปัจจุบันได้มีการพัฒนานำกล่องดนตรีมาผลิตเป็นเพลงในรูปแบบของ mp3 โดยทำการตัดเสียงเนื้อร้องออกไปจนเหลือแต่จังหวะ และทำนอง

2.1.6 เพลงสตริง

เพลงไทยสากล หรือเรียกอีกอย่างว่าเพลงสตริงเป็นเพลงที่บรรเลงโดยวงดนตรีสากลประเภทวงสตริง ส่วนมากจะเป็นเนื้อร้องเนื้อหาเกี่ยวกับสถานะการณ์ปัจจุบัน คำศัพย์จะเกี่ยวกับภาษาที่ได้รับความ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นิยมในเวลานั้น ๆ ดังนั้นเพลงสตริงจึงเป็นเพลงที่ค่อนข้างได้รับความนิยมในต่ละยุค แต่ละเวลาและโอกาส และจะมีการเปลี่ยนแปลงตามกระแสนิยมอย่างรวดเร็ว จะไม่ค่อยมีเพลงอมตะเหมือนเพลงลูกทุ่งและลูกกรุง

ในส่วนของประเทศไทยได้มีการประกวดเพลงไทยสากลแนวใหม่ชิงถ้วยพระราชทานนั่นคือ เพลงสตริงคอมโบ ใช้เครื่องเป่าผสมกีตาร์เป็นหลัก วงชนะเลิศคือวง ดิอิมพอสซิเบิ้ล ซึ่งเป็นวงที่ได้รับความนิยมอย่างมาก ภายหลังปีพุทธศักราช 2521 เพลงสตริงได้รับความนิยมเป็นอย่างมาก เกิดวงดนตรีใหม่ๆ ขึ้นมาเป็นจำนวนมาก นอกจากนี้เพลงสตริงยังเป็นที่นิยมในตลาดจนถึงปัจจุบัน ศิลปินหรือวงดนตรีที่เป็นวงสตริง เช่น วงโปเตโต้,วงบิ๊กแอส, วงบอดี้สแลม, วงแทททูคัลเลอร์ เป็นต้น

2.2 งานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้มีต้นแบบมาจากงานวิจัยของ George Tzanetakis และ Perry Cook [1] ซึ่งเป็นคนกลุ่มแรกๆ ที่ค้นคว้าและวิจัยเกี่ยวกับการจัดหมวดหมู่แนวดนตรี โดยใช้เทคนิค timbral texture, rhythmic content และ pitch content ในการแยกคุณสมบัติของแต่ละประเภทแนวดนตรี และใช้ตัวจำแนกประเภทแบบ K-Nearest Neighbor ซึ่งงานวิจัยที่กล่าวมามีค่าความถูกต้องประมาณ 62%

ในงานวิจัยนี้เราจึงได้ปรับปรุงการทำนายผลของการจัดหมวดหมู่แนวดนตรี โดยทำการเพิ่มฟีเจอร์ให้มีมากขึ้นจากงานวิจัยเดิมใช้ 30 ฟีเจอร์เป็น 43 ฟีเจอร์ และใช้เทคนิค Stacking มาช่วยในการปรับปรุงการทำนายให้มีค่าความถูกต้องมากขึ้น

2.3 การแยกคุณลักษณะ (feature extraction)

การแยกคุณลักษณะเกี่ยวข้องกับการลดปริมาณของจำนวนข้อมูลที่ต้องการอธิบายชุดข้อมูลขนาดใหญ่ เมื่อทำการวิเคราะห์ข้อมูลที่ซับซ้อน หนึ่งในปัญหาที่สำคัญเกิดจากจำนวนตัวแปรที่เกี่ยวข้อง การวิเคราะห์ด้วยตัวแปรจำนวนมากโดยทั่วไปต้องการหน่วยความจำในการคำนวณมาก อาจทำให้อัลกอริทึมการจำแนกประเภทเกิด overfit กับข้อมูลการเทรน และทำให้คุณภาพของข้อมูลใหม่ต่ำลง การแยกคุณลักษณะเป็นศัพท์ทั่วไปสำหรับวิธีการสร้างชุดค่าผสมของตัวแปรเพื่อแก้ไขปัญหาเหล่านี้โดยยังคงอธิบายข้อมูลด้วยความแม่นยำที่เพียงพอ

ในงานวิจัยนี้เราใช้ชุดคุณสมบัติหลักสามประการสำหรับการแยกฟีเจอร์ โดยใช้ชุดฟีเจอร์ดั้งเดิมที่เสนอโดย Tzanetakis & Cook [1] ฟีเจอร์ที่ใช้สามารถแบ่งออกเป็นสามกลุ่ม ได้แก่ จังหวะ (Beat related) , เนื้อเสียง (Timbral texture) และ ระดับเสียง (Pitch related)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.1 Rhythmic Content Features (Beat)

คุณสมบัติเนื้อหาจังหวะเป็นลักษณะการเคลื่อนไหวของสัญญาณเพลงเมื่อเวลาผ่านไปช่วงนึงโดยมีข้อมูลของสัญญาณอยู่ด้วย เช่น ความสอดคล้องของจังหวะ โครงสร้างจังหวะถูกตรวจจับตามระยะเวลาที่เด่นชัดที่สุดของสัญญาณ บีทเป็นฟีเจอร์ที่คำนวณโดยใช้ฮิสโตแกรมคำนวณค่า (Beat Histogram) [1] ซึ่งฮิตโตแกรมสามารถสร้างได้จากการแปลงสัญญาณเวฟเลต โดยแยกสัญญาณออกเป็นความถี่แบบ octave bands เมื่อแยกความถี่ออกแล้วนำไปผ่านตัวกรองความถี่ต่ำแล้วนำสัญญาณมาทำการลดข้อมูล (down sampling) จากนั้นนำมาลบค่าเฉลี่ยของสัญญาณในแต่ละแบนด์ เราเรียกขั้นตอนการทำนี้ว่าการแยกกรอบสัญญาณ (Envelope Extraction) จากนั้นนำสัญญาณที่ผ่านการทำ Envelope มารวมกันโดยวิธี Autocorrelation แล้วทำการเลือกค่าสูงสุดของแอมพลิจูดที่โดดเด่นแล้วนำมาทำเป็นฮิสโตแกรม

2.3.2 Timbral Texture Features

โดยทั่วไปฟีเจอร์ Timbral จะใช้ทางเสียงพูด และการรู้จำเสียงพูด โดยแยกแยะความแตกต่างของเสียงด้วยเนื้อหาที่มีจังหวะเหมือนหรือคล้ายกัน [2] ในการแยกคุณสมบัติของเนื้อเสียงจะแบ่งสัญญาณเสียงออกเป็นช่วงเวลาเล็กๆ หาโดยการนำสัญญาณมาแปลงเป็นฟูริเยร์ในช่วงเวลาอันสั้น (Shot Time Fourier Transform) แล้วนำค่าที่ได้มาหาสเปกตรัมโดยค่าการแปลงฟูริเยร์ในช่วงเวลาอันสั้นยกกำลังสองจะได้เป็นค่าสเปกตรัม

2.3.2.1 สเปกทรัลเซนทรอยด์ (Spectral Centroid)

คุณสมบัติดังกล่าวสะท้อนความเข้มของสัญญาณเสียงและคำนวณเป็นจุดสมดุลของสเปกตรัม โดยนำค่าความถี่ของเสียงมาคำนวณ จะได้ค่าสเปกทรัลเซนทรอยด์ ดังสมการที่ 2.1

$$C = \frac{\sum_{n=1}^{N} M_t[n] * n}{\sum_{n=1}^{N} M_t[n]} \tag{2.1}$$

โดยที่ $M_t[n]$ คือ ค่าแม็กนิจูดของ Fourier transform ที่เฟรม t และความถี่ที่ bin n

2.3.2.2 สเปกทรัลโรลออฟ (Spectral Rolloff)

สเปกทรัล roll off คืออัตราการลดทอนของสัญญาณต่อจำนวนความถี่ที่เปลี่ยนไป กำหนดความถี่ R = R (r) ที่สอดคล้องกับ r% ของการกระจายขนาดแม็กนิจูด ในงานวิจัยนี้ใช้ค่า rolloff r = 80%

$$\sum_{n=1}^{R_t} M_t[n] = r * \sum_{n=1}^{N} M_t[n] \tag{2.2}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.2.3 สเปกทรัลฟลักซ์ (Spectral Flux)

สเปกทรัลฟลักซ์ถูกกำหนดให้เป็นวิวัฒนาการของความถี่พร้อมเวลาและถูกคำนวณเป็นความแตกต่างของขนาดแม็กนิจูดระหว่างเฟรมปัจจุบันกับเฟรมก่อนหน้า สเปกตรัมฟลักซ์เป็นการวัดปริมาณการเปลี่ยนแปลงสเปกตรัมเฉพาะที่ (local spectral)

$$F = \sum_{n=1}^{N}(N_t[n] - N_{t-1}[n])^2 \quad (2.3)$$

โดยที่ $N_t[n]$ and $N_{t-1}[n]$ คือ การนอมอไรซ์แม็กนิจูดของ Fourier transform ที่หน้าต่าง t และ $t-1$ ตามลำดับ

2.3.2.4 ซีโร่ครอสซิ่ง (Zero Crossings)

ซีโร่ครอสซิ่งเกิดขึ้นเมื่อตัวอย่างทดแทนในสัญญาณดิจิตอลมีสัญญาณต่างกัน ฟีเจอร์นี้ถูกกำหนดให้อยู่ในโดเมนของเวลา ซึ่งมีประโยชน์ในการตรวจสอบปริมาณของเสียงรบกวน (noise) ในสัญญาณ

$$\sum_{n=1}^{N}|sign(x[n]) - sign(x[n-1])| \quad (2.4)$$

โดยที่ฟังก์ชัน $sign$ เป็น 1 สำหรับการขัดแย้งเชิงบวกและ 0 สำหรับการขัดแย้งเชิงลบ

2.3.2.5 ความเข้มเสียง

ความเข้มของเสียงสามารถคำนวณได้จากสูตร Root mean square (RMS) เพื่อใช้ในการจำแนกประเภทของดนตรีโดยดูจากความเข้มของเสียง ความเข้มของเสียงมากจะเป็นแนวดนตรีที่มีจังหวะเร็ว แต่ถ้าความเข้มเสียงน้อยจะเป็นแนวดนตรีที่มีจังหวะช้า

$$v(n) = \sqrt{\frac{1}{N}\sum_{t=0}^{N-1} S_n^2(i)} \quad (2.5)$$

โดย v(n) คือ ค่าความเข้มเสียง

N คือ ขนาดของเฟรม

$S_n(i)$ คือ สุ่มตัวอย่างลำดับที่ i ของเฟรมที่ n

n คือ ลำดับของเฟรม

2.3.3 ระดับเนื้อเสียง (Pitch Content Features)

อัลกอรึทึมการตรวจจับระดับเสียง (pitch) [4] อธิบายโดย Tolonen และ Karjalainen [3] Pitch เป็นความถี่พื้นฐานของสัญญาณเสียงพูด พื้นที่การเน้นเสียงส่วนใหญ่จะประเมินโดยพิจารณาคุณสมบัติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ของระดับเสียง การศึกษาเหล่านี้พิจารณา ความถี่ pitch, 1การวิเคราะห์ทางสถิติของค่าเฉลี่ยของระดับเสียง และส่วนเบี่ยงเบนมาตรฐาน สัญญาณ pitch ถูกสร้างขึ้นเนื่องจากการสั่นสะเทือนของเสียงร้อง

ในการแปลงสัญญาณเสียงให้อยู่ในโดเมนของความถี่เราอาศัยเทคนิคสัมประสิทธิ์เซปสตรัมบนแกนความถี่เมล (Mel-Frequency Cepstral Coefficients, MFCC) [1] และการแปลงฟูริเยร์ช่วงเวลาสั้น (Short Time Fourier Transform) MFCC เป็นลักษณะเด่นที่ปรับปรุงมาจากสัมประสิทธิ์เซปสตรัมปกติด้วยการปรับสเกลขอลสเปกตรัมให้อยู่บนสเกลที่เหมาะสมสำหรับการรับฟังของมนุษย์ จึงได้มีการออกแบบสเกลของสเปกตรัมให้สามารถเก็บรายละเอียดของสัญญาณเสียงช่วงความถี่ต่ำได้มากกว่า เรียกว่า สเกลเมล (Mel scale) สูตรในการแปลงหน่วยจากเฮิรตซ์ไปเป็นหน่วยของสเกลเมล ดังสมการที่ 2.6

$$m = 2595log_{10}(\frac{f}{700} + 1) \quad (2.6)$$

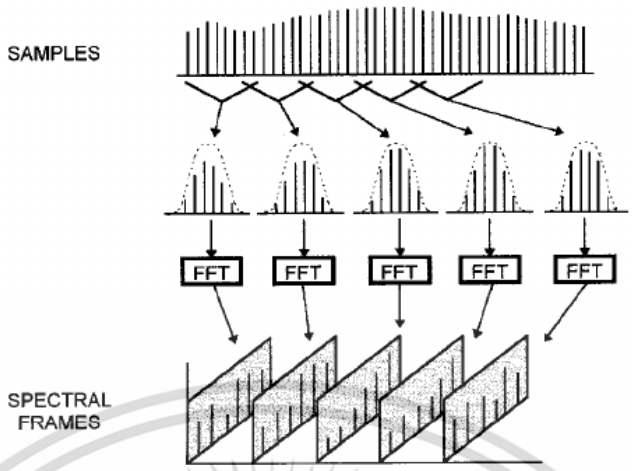
โดยที่ m คือ ความถี่ในหน่วยของสเกลเมล และ f คือความถี่ของเสียงในแนวเส้นตรง นำสัญญาณที่ได้จากการเลื่อนหน้าต่างไปผ่านกระบวนการแปลงฟูรีเยช่วงเวลาสั้น (Short Time Fourier Transform) เพื่อเปลี่ยนสัญญาณเสียงให้อยู่ในโดเมนความถี่

2.3.4 การแปลงฟูริเยร์ช่วงเวลาสั้น (Short Time Fourier Transform)

เป็นกระบวนการทางคณิตศาสตร์ที่ใช้อธิบายโครงสร้างของระบบสัญญาณที่ประกอบด้วยสัญญาณเฉพาะหลายๆ สัญญาณมารวมกัน มีหลักการคือแปลงสัญญาณเป็นโดเมนความถี่ (frequency domain)

การวิเคราะห์สัญญาณในโดเมนเวลาและความถี่มีหลายวิธีด้วยกันเช่น การวิเคราะห์สัญญาณโดยการแปลงฟูริเยร์ในช่วงเวลาอันสั้น (Shot Time Fourier Transform) เป็นวิธีการหนึ่งสำหรับการแสดงสัญญาณในโดเมนเวลาและความถี่โดยอาศัยเทคนิค Fourier Transform วิธีการทำงานโดยจะเลือกบริเวณของสัญญาณที่ต้องการวิเคราะห์หลังจากนั้นจะทำการคำนวณ Fourier Transform ของสัญญาณที่อยู่ใน window ทำให้ได้ข้อมูลความถี่ที่เวลานั้นๆ ต่อมา เลื่อน window function ไปที่บริเวณอื่นของสัญญาณ แล้วทำการคำนวณ Fourier Transform ทำเช่นนี้ไปเรื่อยๆ จนครบสัญญาณ สุดท้ายจะได้ข้อมูลของสัญญาณในโดเมนเวลาและความถี่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.1 Shot Time Fourier Transform

2.3.5 การแปลงเวฟเลต (Wavelet Transform)

การวิเคราะห์สัญญาณด้วยวิธีการแปลงเวฟเลตได้รับการพัฒนาเพื่อใช้ในการแก้ปัญหาที่ไม่สามารถดำเนินการได้ในการแปลงฟูริเยร์และการแปลงฟูริเยร์ในช่วงเวลาอันสั้น โดยสามารถปรับช่วงเวลาและความถี่ให้เหมาะสมตามความต้องการได้ หากขยายช่วงเวลาของฟังก์ชันหน้าต่างให้ยาวขึ้นส่งผลให้ข้อมูลช่วงความถี่ต่ำมีความเที่ยงตรง และถ้าลดช่วงเวลาของฟังก์ชันหน้าต่างให้น้อยลงจะส่งผลให้ข้อมูลช่วงความถี่สูงมีความเที่ยงตรงเช่นกัน ดังแสดงในรูปที่ 2.2

รูปที่ 2.2 การวิเคราะห์สัญญาณด้วยการแปลงเวฟเลต

ซึ่งการวิเคราะห์สัญญาณด้วยวิธีการแปลงเวฟเลตจะแสดงข้อมูลเชิงเวลาและขนาด สามารถใช้การวิเคราะห์สัญญาณแบบเฉพาะที่ได้ โดยเวฟเลตมีลักษณะพิเศษที่สำคัญคือ สมาชิกของฟังก์ชันเวฟเลตเป็นสัญญาณที่เกิดจากต้นแบบอันเดียวกัน และต้นแบบดังกล่าวนี้เรียกว่า เวฟเลตแม่ (mother wavelet) ซึ่งคลื่นเวฟเลตในแต่ละกลุ่มเกิดจากการปรับสเกลที่แสดงการยืดหรือหดของเวฟเลตแม่ และการเลื่อนตำแหน่งที่แสดงบนแกนเวลา ดังสมการที่ 2.7

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{2.7}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สัญญาณที่ผ่านกระบวนการแปลงเวฟเลตจะมีการแตกสัญญาณในรูปของเวฟเลตแม่โดยมีการปรับสเกลและตำแหน่งที่แตกต่างกันออกไป ซึ่งสามารถแบ่งรูปแบบการแปลงเวฟเลตทั่วไปได้ 2 ประเภทคือ การแปลงเวฟเลตแบบต่อเนื่อง (continuous wavelet transform) และการแปลงเวฟเลตแบบเต็มหน่วย (discrete wavelet transform)

1.การแปลงเวฟเลตแบบต่อเนื่อง (continuous wavelet transform) ได้รับการพัฒนาเพื่อแก้ไขปัญหาของการแปลงฟูริเยร์ในช่วงเวลาอันสั้น มีความคล้ายคลึงกันในส่วนของสัญญาณเดิมจะถูกคูณด้วยฟังก์ชันเวฟเลต ซึ่งการแปลงฟูริเยร์ในช่วงเวลาอันสั้นนั้นจะถูกคูณด้วยฟังก์ชันหน้าต่าง โดยความกว้างของหน้าต่างในการแปลงเวฟเลตจะเปลี่ยนแปลงไปเรื่อยๆ ตลอดการวิเคราะห์สัญญาณในทุกๆ ค่าของความถี่

2.การแปลงเวฟเลตแบบเต็มหน่วย (discrete wavelet transform) พัฒนาขึ้นเพื่อแก้ไขปัญหาของการแปลงเวฟเลตแบบต่อเนื่อง โดยมีรูปแบบการปรับสเกลและการเลื่อนตำแหน่งในลักษณะเป็นช่วงๆ ที่ไม่ต่อเนื่องกัน

2.4 การคัดเลือก feature (feature selection)

ในการจัดหมวดหมู่คุณสมบัติที่ใช้ในการอธิบายตัวอย่างในชุดข้อมูล เนื่องจากฟีเจอร์ที่มีประโยชน์มักจะไม่รู้ล่วงหน้าจึงมีฟีเจอร์จำนวนมากเพื่อใช้อธิบายตัวอย่างในชุดข้อมูล เห็นได้ชัดว่าไม่ใช่ฟีเจอร์ทั้งหมดที่มีประโยชน์สำหรับการจัดหมวดหมู่ ฟีเจอร์ที่ไม่เกี่ยวข้องหรือซ้ำซ้อนอาจลดประสิทธิภาพการจัดหมวดหมู่ [5] อัลกอริทึมการจำแนกประเภทมักประสบปัญหาจากมิติของข้อมูลเนื่องจากฟีเจอร์จำนวนมากในชุดข้อมูล ดังนั้นการเลือกฟีเจอร์หรือที่เรียกว่าการเลือกตัวแปรจะถูกเสนอเป็นขั้นตอนการประมวลผลข้อมูลล่วงหน้าเพื่อลดหรือกำจัดฟีเจอร์ที่ไม่เกี่ยวข้องและซ้ำซ้อน

การเลือกฟีเจอร์ถูกกำหนดโดยนักวิจัยหลายคนจากมุมมองที่แตกต่างกัน แต่ส่วนใหญ่มีความคล้ายคลึงกัน [5] นิยามต่อไปนี้เป็นคำจำกัดความที่ครอบคลุมของการเลือกฟีเจอร์

- การปรับปรุงความถูกต้องของการจำแนกประเภท: การเลือกฟีเจอร์ คือการเลือกชุดย่อยของฟีเจอร์สำหรับการปรับปรุงประสิทธิภาพการจัดหมวดหมู่ หรือลดความซับซ้อนของแบบจำลองโดยไม่ลดความแม่นยำในการจำแนกประเภทของตัวแยกประเภทที่สร้างขึ้น

ความเหมาะสม: การเลือกฟีเจอร์ คือการค้นหาเซตย่อยขนาดย่อขนาดเล็กที่สุดที่จำเป็นและเพียงพอที่จะอธิบายข้อมูล [10]

Feature selection สามารถแบ่งออกได้เป็น 3 วิธี ได้แก่ วิธีฟิลเตอร์ (Filter Method), วิธีแรปเปอร์ (Wrapper Method) และวิธีฝังตัว (Embedded Method)

1.วิธี Filter เป็นการคัดเลือกฟีเจอร์โดยใช้การคำนวณหาค่าน้ำหนักซึ่งอาจจะเป็นค่าความสัมพันธ์ระหว่างแต่ละฟีเจอร์และคลาสต่างๆ และจะเลือกฟีเจอร์โดยเรียงลำดับตามค่าน้ำหนักที่คำนวณได้แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เลือกฟีเจอร์ที่มีค่าน้ำหนักมากกว่าที่ต้องการมาใช้งานต่อไป วิธีการนี้ต่างจากวิธีการ Wrapper ตรงที่วิธีการนี้จะไม่มีการสร้างโมเดลเพื่อคัดเลือกฟีเจอร์ เทคนิคในการคำนวณค่าน้ำหนักของฟีเจอร์ต่างๆ มีหลายวิธี เช่น Information Gain, Chi-Square หรือ Correlation

2.วิธี Wrapper [7] เป็นการคัดเลือกฟีเจอร์ด้วยการสร้างโมเดล (classification model) ขึ้นมาจากเซตของฟีเจอร์ที่กำหนดไว้และวัดประสิทธิภาพการทำงานของโมเดล และเลือกเซตของฟีเจอร์ที่ทำให้โมเดลมีประสิทธิภาพมากที่สุดมาใช้งาน เช่น โมเดลที่ให้ค่าความถูกต้อง (accuracy) มากที่สุด การคัดเลือกฟีเจอร์ด้วยวิธีการนี้แบ่งย่อยได้เป็น 2 แบบใหญ่ๆ คือ Forward Selection เป็นการสร้างโมเดลโดยการเพิ่มฟีเจอร์ทีละ 1 ฟีเจอร์ ถ้าฟีเจอร์ที่ใส่เพิ่มให้ประสิทธิภาพที่ดีก็จะเก็บไว้และเลือกฟีเจอร์อื่นๆ มาเพิ่มต่อไปจนประสิทธิภาพของโมเดลไม่ได้ดีขึ้นก็จะหยุดทำงาน
Backward Elimination เป็นการสร้างโมเดลที่เริ่มจากการใช้ฟีเจอร์ทั้งหมดก่อนและตัด (eliminate) ฟีเจอร์ที่ไม่สำคัญทิ้งไปทีละฟีเจอร์ถ้าประสิทธิภาพดีขึ้นก็ตัดฟีเจอร์อื่นๆ ต่อไป

3.วิธี Embedded เป็นวิธีที่การคัดเลือกตัวแปรเป็นส่วนหนึ่งในกระบวนการจำแนกประเภทด้วย โดยทำการเลือกเซตย่อยที่มีตัวแปรที่เหมาะสมในแต่ละขั้นตอนวิธี พร้อมไปกับการสร้างตัวแบบสำหรับการจำแนกประเภท ซึ่งวิธีฝังตัวมีลักษณะคล้ายกับวิธีแรปเปอร์ที่ว่ามีการผูกติดกับวิธีการจำแนกประเภทที่เฉพาะเจาะจง แต่ใช้เวลาในการประมวลผลน้อยกว่าวิธีแรปเปอร์

หลายๆ ครั้งในการจำแนกประเภทข้อมูล (classification) จะพบว่าแอตทริบิวต์ (attribute) หรือฟีเจอร์ (feature) มีจำนวนมาก เช่น การจำแนกประเภทข้อความทัศนคติ (sentiment) ออกเป็นเชิงบวก (positive) หรือเชิงลบ (negative) นั้นจะมีจำนวนคำในข้อความต่างๆ ที่ใช้เป็นฟีเจอร์จำนวนมาก ฟีเจอร์เหล่านี้บางอันก็ไม่ได้มีความสำคัญในการแบ่งแยกคลาส (class) ออกเป็นเชิงบวกหรือเชิงลบได้ ดังนั้นจึงจำเป็นต้องทำการคัดเลือกฟีเจอร์ที่สำคัญมาใช้งาน ในงานวิจัยนี้เลือกใช้วิธีการของ Particle Swarm Optimization (PSO)

2.4.1 การหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization)

การหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization) [8] หรือ PSO พัฒนาขึ้นโดยอีเบอร์ฮาร์ดและเคนเนดี้ (J.Kennedy and R.Eberthart, 1995) โดยมีแนวคิดมาจากการจำลองวิธีการหาค่าเหมาะสมที่สุดจากพฤติกรรมการเดินทางหรือหาอาหารของฝูงสัตว์ โดยเฉพาะฝูงนกฝูงปลา

ตัวอย่างเช่น มีนกฝูงหนึ่งกำลังบินสุ่มหาอาหารอยู่พื้นที่หนึ่ง ที่ซึ่งมีอาหารอยู่เพียงชิ้นเดียวในบริเวณนั้น นกทุกตัวไม่รู้ว่าอาหารอยู่ที่ตำแหน่งไหน แต่สัณชาตญาณพวกมันรู้ว่ามันอยู่ห่างจากอาหารเท่าไหร่ โดยเมื่อพวกมันบินไปในแต่ละช่วง พวกมันจะหาระยะห่างของมันจากแหล่งอาหาร และนกทั้งฝูงจะเลือก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บินตามนกตัวที่อยู่ใกล้อาหารมากที่สุด จากนั้นพวกมันก็จะบินช่วงต่อไป โดยที่นกทั้งฝูงจะทำซ้ำเช่นนี้ไปเรื่อยๆ จนพวกมันบินมาถึงแหล่งอาหาร จากวิธีการหาอาหารของฝูงนกเราจะเห็นว่า ในการกำหนดตำแหน่งระยะห่างของตัวมันจากแหล่งอาหาร และการตัดสินใจบินตามนกตัวที่บินอยู่ใกล้แหล่งอาหารมากที่สุด ในแต่ละช่วงจะใช้เวลาน้อยมาก จนเราดูเหมือนมันบินเกาะกลุ่มกันไปแบบต่อเนื่อง

PSO เป็นการจำลองการหาอาหารของฝูงนก ในการหาค่าเหมาะที่สุดแบบนี้ นกแต่ละตัวในฝูงจะถูกแทนด้วยอนุภาค (particle) ในอนุภาคแต่ละตัวจะมีค่า ฟิตเน็ส (fitness value) ที่บอกถึงระยะห่างของตัวมันจากแหล่งอาหาร โดยอนุภาคทั้งหมดจะบินตามอนุภาคที่มีค่าฟิตเน็ตที่ดีที่สุดในแต่ละช่วง/รอบ (Iteration)

แนวคิดของ PSO จะเริ่มต้นด้วยการสุ่มหาตำแหน่งของอนุภาค (ซึ่งตำแหน่งต่างๆ ของอนุภาคเหล่านั้นก็คือคำตอบที่เป็นไปได้) ขึ้นมาชุดหนึ่ง จากนั้นก็จะหาค่าเหมาะสมที่สุดด้วยการปรับปรุงค่าในแต่ละรอบของการตัดสินใจ โดยที่อนุภาคแต่ละตัวจะมีการปรับปรุงค่าด้วยการเปลี่ยนตำแหน่งตามค่าที่ดีที่สุด (Best) ซึ่งมี 2 ค่า ได้แก่

- ค่าที่ดีที่สุดของอนุภาค (pbest : particle best) คือค่าตำแหน่งที่ดีที่สุดของการเคลื่อนที่ที่ผ่านมาของตัวมันเอง
- ค่าที่ดีที่สุดของสากล (gbest : global best) คือค่าที่เป็นค่าที่ดีที่สุดการเคลื่อนที่ที่ผ่านมาของทั้งกลุ่ม (ฝูง)

การทำงานของ PSO เป็นกระบวนการที่ทำงานเป็นรอบ (Iteration) ซึ่งในแต่ละรอบของการทำงานความเร็วของอนุภาคแต่ละตัวจะถูกปรับปรุงโดยมีตัวแปรที่สำคัญ 3 ตัว คือ

- ความเร็วปัจจุบันของอนุภาคนั้น (velocity)
- ข้อมูลที่อนุภาคมีอยู่ (pbest)
- ข้อมูลรวมของอนุภาคทั้งกลุ่ม (gbest)

หลังจากนั้นอนุภาคแต่ละตัวจะปรับตำแหน่งของมันโดยใช้ความเร็วใหม่ที่คำนวณได้ โดยพิจารณาจากสมการดังต่อไปนี้

- คำนวนค่าความเร็วใหม่ของอนุภาคตามสมการ 2.8

$$V_i(t+1) = wv_i(t) + c_1 r_1\big(p_i(t) - x_i(t)\big) + c_2 r_2(p_g(t) - x_i(t)) \quad (2.8)$$

- ปรับปรุงค่าตำแหน่งใหม่ของอนุภาคตามสมการ 2.9

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (2.9)$$

โดยที่ c_1,c_2 คือค่าคงที่ บางครั้งถูกเรียกว่า ค่าแฟคเตอร์การเรียนรู้ (เอาไว้ถ่วงน้ำหนักความจำ) โดยทั่วไปจะเท่ากับ 2 แต่ก็สามารถใช้ค่าอื่นๆ ได้ โดยที่จะมีช่วงระหว่าง [0,4] หรือ 0 ถึง 4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สมการ fitness ใช้เพื่อหาค่าที่เหมาะสมที่สุด ในงานวิจัยนี้ใช้ตัวจำแนกประเภทนำมาทำเป็นสมการฟังก์ชั่น fitness โดยการคำนวณค่าความถูกต้องของแต่ละตัวแยกประเภทเพื่อดูว่าตัวแยกประเภทชนิดใดให้ค่า fitness ได้ดีที่สุด หรือก็คือให้ค่าความถูกต้องได้สูงที่สุดโดยที่มีจำนวนประชากร และจำนวนการเลือกฟีเจอร์ที่เหมือนกัน

2.5 วิธีการแยกประเภท (Classification)

การทำเหมืองข้อมูล (Data Mining) คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น[5] ซึ่งเป็นส่วนช่วยในการจัดเก็บและตีความหมายข้อมูลจากเดิมที่มีการจัดเก็บข้อมูลอย่างง่ายๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้มีเทคนิคที่นิยมใช้ คือ การจำแนกประเภทข้อมูล (Classification) ซึ่งเป็นกระบวนการสร้างโมเดลเพื่อทำนายกลุ่มของข้อมูลใหม่โดยข้อมูลทั้งหมดจะมีการแบ่งออกเป็น 2 กลุ่มคือกลุ่มข้อมูลสอน (Training Data) เป็นชุดข้อมูลที่มีบทบาทในการสร้างโมเดลจำแนกประเภทข้อมูลขึ้นมาและกลุ่มข้อมูลทดสอบ (Test Data) เป็นชุดข้อมูลประเมินความถูกต้องของโมเดลโดยกระบวนการสร้างตัวโมเดลจำแนกประเภทข้อมูลแบ่งออกเป็น 3 ขั้นตอนดังนี้

1. การสร้างโมเดลการจำแนกประเภทเป็นขั้นตอนที่อาศัยการเรียนรู้จากข้อมูลสอนที่ได้กำหนดประเภทไว้เรียบร้อยแล้ว
2. การประเมินผลโมเดลเป็นขั้นตอนตรวจสอบความถูกต้องโดยใช้ข้อมูลทดสอบ และปรับปรุงโมเดลจำแนกประเภทจนกว่าจะได้ค่าความถูกต้องในระดับที่ยอมรับได้
3. การนำโมเดลไปใช้งาน เป็นขั้นตอนการนำโมเดลจำแนกประเภทที่สร้างขึ้นมาใช้กับข้อมูลที่ไม่เคยเห็นมาก่อนเพื่อทำนายและกำหนดกลุ่มให้กับข้อมูลนั้น

Classification เป็นการจัดประเภทข้อมูลที่มีลักษณะเดียวกันให้อยู่ในกลุ่มเดียวกัน เป็นหนึ่งในวิธีการเรียนรู้แบบมีหรือไม่มีผู้สอน (Supervised Learning) ที่เราต้องเตรียมตัวอย่าง (Training Data) ให้กับระบบก่อน โดยแต่ละตัวอย่างของเราจะมีป้ายกำกับ (Label) เพื่อบอกว่าตัวอย่างแต่ละตัวของเราเป็นประเภทใด ซึ่งในงานวิจัยนี้ใช้ตัวแยกประเภทที่เป็นที่นิยมและนำมาใช้งานได้ง่ายจำนวน 6 ชนิด คือ K-nearest neighbor (KNN), Decision Tree classifiers (DT), Random Forest, Support Vector Machine (SVM), Naive Bayes และ Neural Classifiers

2.5.1 K-nearest neighbor (kNN)

K-nearest neighbor [9] เป็นประเภทของอัลกอริทึมการเรียนรู้ตามตัวอย่าง เมื่อใช้ kNN สำหรับการจัดหมวดหมู่มันจะคำนวณระยะทางระหว่างตัวอย่างในชุดทดสอบ และทุกตัวอย่างในชุดการเทรน kNN กำหนดตัวอย่างทดสอบให้กับคลาสที่พบมากที่สุดในหมู่เพื่อนบ้านที่ใกล้ที่สุดโดยที่ k เป็นจำนวนเต็ม

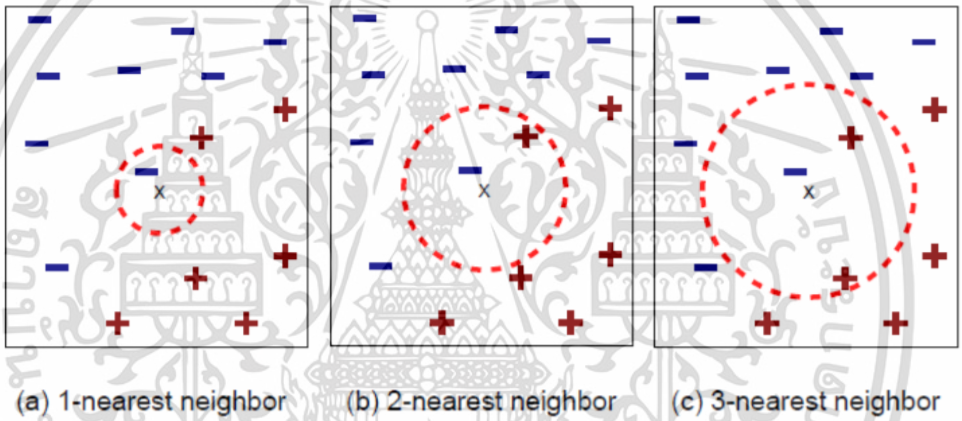
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บวกโดยทั่วไปจะมีขนาดเล็ก หาก k = 1 ตัวอย่างทดสอบจะถูกกำหนดให้กับคลาสของเพื่อนบ้านที่ใกล้เคียงที่สุดตัวเดียวเท่านั้น ข้อดีของวิธีการนี้ คือสามารถให้ผลประสิทธิภาพที่ดีถ้า training data มีขนาดใหญ่

ใช้การวัดระยะห่างแบบยูคลิเดียน (Euclidean Distance) นิยามตัวอย่าง x ด้วยเวกเตอร์คุณลักษณะ

$x = \langle a_1(x), a_2(x), \ldots, a_n(x) \rangle$ และคำนวณระยะห่างระหว่างตัวแปรตัวอย่าง x_i และ x_j ที่แทนด้วย $d(x_i, x_j)$ ได้ดังสมการที่ 2.10

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(x_j))^2} \quad (2.10)$$



รูปที่ 2.3 K-nearest neighbor ที่ค่า K ต่างๆ โดยมีจุด X เป็นตัวเทียบ

2.5.2 Decision Tree classifiers (DT)

การเรียนรู้แบบ DT เป็นอัลกอริทึมในการประมาณฟังก์ชั่นที่มีค่าไม่ต่อเนื่อง [9] วิธีการของ Decision Tree คือเทคนิคที่ให้ผลลัพธ์ในลักษณะเป็นโครงสร้างของต้นไม้ ภายในต้นไม้จะประกอบไปด้วยโหนด (node) ซึ่งแต่ละโหนดจะมีเงื่อนไขของคุณลักษณะเป็นตัวทดสอบ กิ่งของต้นไม้ (branch) แสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือกทดสอบ และใบ (leaf) เป็นสิ่งที่อยู่ล่างสุดของต้นไม้แสดงถึงกลุ่มของข้อมูล (class) ก็คือผลลัพธ์ที่ได้จากการพยากรณ์ ซึ่งข้อดีของวิธีการนี้ คือให้ผลการพยากรณ์ที่แม่นยำ

การสร้างโมเดล decision tree จะทำการคัดเลือกแอตทริบิวต์ที่มีความสัมพันธ์กับคลาสมากที่สุดขึ้นมาเป็นโหนดบนสุดของ tree (root node) หลังจากนั้นก็จะหาแอตทริบิวต์ถัดไปเรื่อยๆ ในการหาความสัมพันธ์ของ

แอตทริบิวต์นี้จะใช้ตัววัด ที่เรียกว่า Information Gain (IG) ค่านี้คำนวณได้จากสมการที่ 2.11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$IG(S,A) = Entropy(S) - \sum_{Values(A)} \frac{|Sv|}{|S|} Entropy(Sv) \quad (2.11)$$

โดย A คือ Attribute A

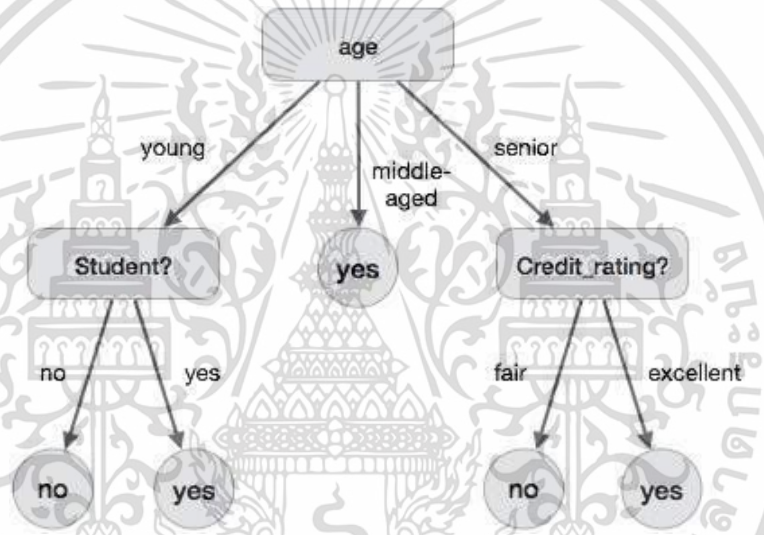
$|Sv|$ คือ สมาชิกของ Attribute A ที่มีค่า v

$|S|$ คือ จำนวนสมาชิกของกลุ่มตัวอย่าง

$$Entropy(S) = \sum_{i=1}^{c} -p_i \, log_2 p_i \quad (2.12)$$

โดย S คือ Attribute ที่นำมาวัดค่า Entropy

p_i คือ สัดส่วนของจำนวนสมาชิกของกลุ่ม i กับจำนวนสมาชิกทั้งหมดของกลุ่มตัวอย่าง

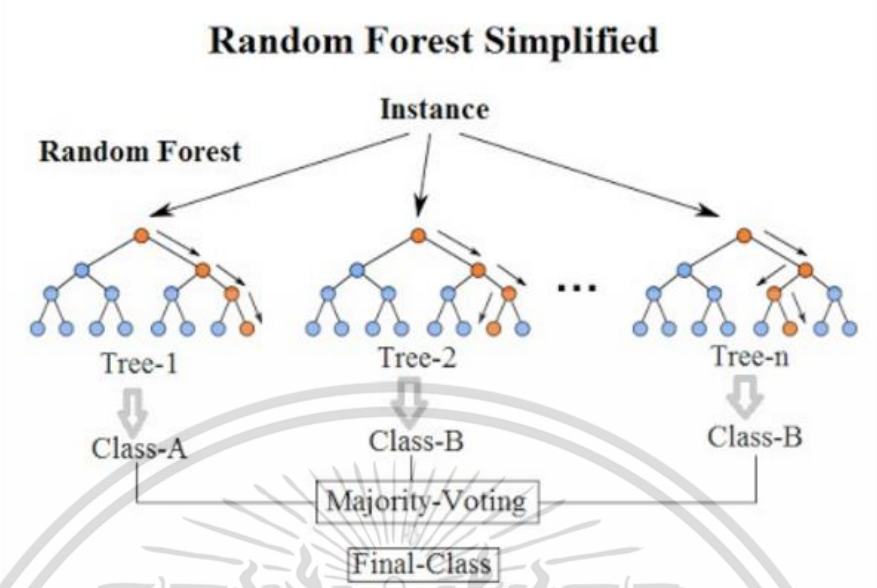


รูปที่ 2.4 แสดงตัวอย่างของการตัดสินใจแบบต้นไม้ (Decision Tree)

2.5.3 Random Forest

อัลกอริทึม Random Forest คือ การสร้างโมเดลด้วยวิธีการ Decision Tree ขึ้นมาหลายๆโมเดล โดยวิธีการสุ่มตัวแปร แล้วนำผลที่ได้แต่ละโมเดลมารวมกันพร้อมนับจำนวนผลที่มีจำนวนซ้ำกันมากที่สุด สกัดออกมาเป็นผลลัพธ์สุดท้าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.5 แสดงตัวอย่างของ Random Forest โดยนำต้นไม้หลายๆ ต้นมาทำการ vote

2.5.4 Support Vector Machine (SVM)

SVM เป็นวิธีการเรียนรู้ที่เป็นที่นิยม แนวคิดหลักของ Support Vector Machine [9] คือการหาระนาบการตัดสินใจในการแบ่งข้อมูลออกเป็นสองส่วน โดยจะพยายามสร้างเส้นแบ่ง (Hyperplane) ที่เป็นเส้นตรงระหว่างกึ่งกลางของกลุ่มข้อมูลให้มีระยะห่างระหว่างขอบเขตของทั้งสองกลุ่มมากที่สุด โดยเน้นไปยังเส้นแบ่งแยกกลุ่มข้อมูลได้ดีที่สุด (optimal separating hyperplane) เมื่อเราพิจารณาข้อมูลที่ประกอบด้วยข้อมูล 2 กลุ่มดังสมการที่ 2.13

$$D = \{(x_i, y_i); i = 1,2, \dots, n\} \tag{2.13}$$

เมื่อ $x_i = (x_{i1}, x_{i2}, \dots, x_{im}) \epsilon R^m$

$y_i \epsilon \{1, -1\}$ โดย 1 คือข้อมูลกล่ม 1 และ -1 คือข้อมูลกลุ่ม 2

ซึ่งเป็นการกำหนดกลุ่มเป้าหมายให้ SVM โดยที่ SVM นั้นมุ่งเป้าเพื่อหาฟังก์ชันการตัดสินใจที่สามารถแบ่งแยกค่าที่ไม่ทราบได้ดังสมการที่ 2.14

$$f(x) = sign\left\{\sum_{k=1}^{n_v} w_k \varphi_k(x) \varphi_k(x_k) + b\right\} \tag{2.14}$$

แต่เดิมการแยกคลาสไบนารี (k = 2) มักต้องเลือกทำมากกว่าสองครั้ง เราจึงแก้ปัญหาการจำแนกประเภทหลายคลาส (k> 2) ด้วยการใช้วิธีการของ Error Correcting Output Codes (ECOC) [13] เข้ามาช่วยแก้ปัญหา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.5 Naive Bayes

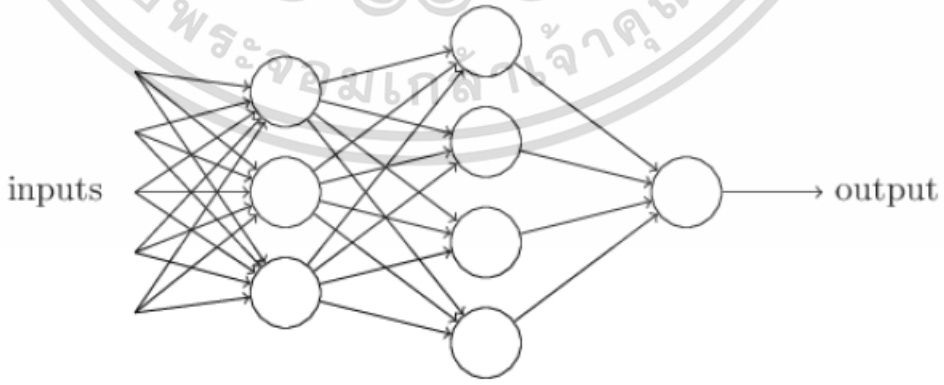
ตัวจำแนกประเภทแบบเบย์อย่างง่ายคือโมเดลการจำแนกประเภทข้อมูลที่ใช้หลักความน่าจะเป็นซึ่งอยู่บนพื้นฐานของ Bayes' Theorem [12] และสมมติฐานที่ให้การเกิดของเหตุการณ์ต่าง ๆ เป็นอิสระต่อกัน ถ้ากำหนดให้ P(h) ความน่าจะเป็นที่จะเกิดเหตุการณ์ h และ P(h|D) คือความน่าจะเป็นที่จะเกิดเหตุการณ์ h เมื่อเกิดเหตุการณ์ D จากตัวแปรที่กำหนด เราสามารถทำนายเหตุการณ์ที่พิจารณาจากเหตุการณ์ต่าง ๆ ได้ดังสมการที่ 2.15

$$P(h \mid D) = P(D \mid h) * P(h) / P(D) \quad (2.15)$$

2.5.6 Neural Network

โครงข่ายประสาทเทียม หรือ ข่ายงานประสาทเทียม (artificial neural network) คือ โมเดลทางคณิตศาสตร์หรือโมเดลทางคอมพิวเตอร์สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนคชันนิสต์ (connectionist) แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ (bioelectric network) ในสมอง ซึ่งประกอบด้วย เซลล์ประสาท (neurons) และ จุดประสานประสาท (synapses) ตามโมเดลนี้ ข่ายงานประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาท จนเป็นเครือข่ายที่ทำงานร่วมกัน

นักวิจัยส่วนใหญ่ในปัจจุบันเห็นตรงกันว่าข่ายงานประสาทเทียมมีโครงสร้างแตกต่างจากข่ายงานในสมอง แต่ก็ยังเหมือนสมอง ในแง่ที่ว่าข่ายงานประสาทเทียม คือการรวมกลุ่มแบบขนานของหน่วยประมวลผลย่อยๆ และการเชื่อมต่อนี้เป็นส่วนสำคัญที่ทำให้เกิดสติปัญญาของข่ายงาน เมื่อพิจารณาขนาดแล้ว สมองมีขนาดใหญ่กว่าข่ายงานประสาทเทียมอย่างมาก รวมทั้งเซลล์ประสาทยังมีความซับซ้อนกว่าหน่วยย่อยของข่ายงาน อย่างไรก็ดีหน้าที่สำคัญของสมอง เช่นการเรียนรู้ ยังคงสามารถถูกจำลองขึ้นอย่างง่ายด้วยข่ายงานประสาท

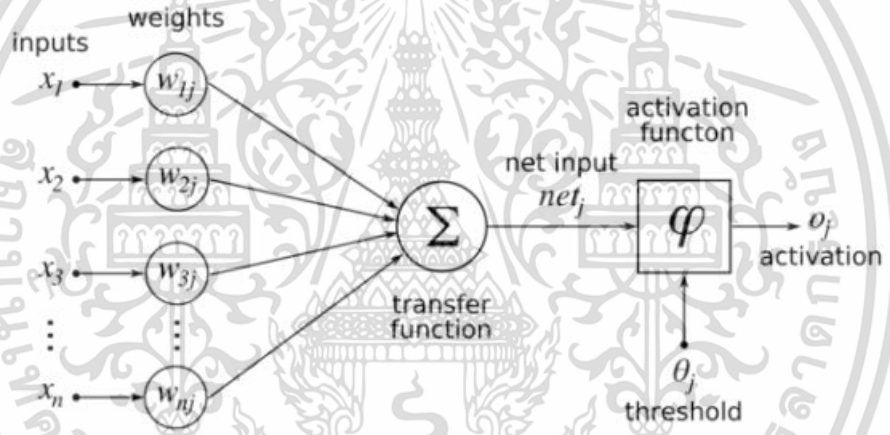


รูปที่ 2.6 รูปแบบพื้นฐาน Neural Network

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.6.1 เซลล์ประสาทเทียม

แบบจำลองของเซลล์ประสาททางชีวภาพ หรือที่เรียกว่าเซลล์ประสาทเทียม (Artificial Neuron) แสดงได้ดังรูปที่ 2.7 โดยประกอบด้วยโครงสร้างพื้นฐานที่สำคัญคือ หน่วย (Node) หรือ ยูนิต (Unit) ตัวแปรด้านเข้า (Input) ตัวแปรด้านออก (Output) และค่าถ่วงน้ำหนัก (Weight) ซึ่งสามารถสรุปความสัมพันธ์ระหว่างเซลล์ประสาททางชีววิทยาและเซลล์ประสาทเทียมดังนี้ การประมวลผล จะเกิดขึ้นในหน่วยประมวลผลย่อยคือ หน่วย (Node) หรือ ยูนิต (Unit) ซึ่งจำลองมาจากลักษณะการทำงานของตัวเซลล์ การส่งสัญญาณระหว่างหน่วยด้วยส่วนที่เชื่อมติดกัน แต่ละการเชื่อมต่อประกอบด้วยค่าน้ำหนักที่แตกต่างกัน โดยขึ้นอยู่กับอิทธิพลที่หน่วยจะได้รับจากหน่วยอื่นๆ โดยค่าน้ำหนักที่ได้รับจะทำหน้าที่เสมือนความรู้ที่ถูกรวบรวม
ไว้ใช้แก้ปัญหาเฉพาะอย่างของมนุษย์ ภายในหน่วยมีฟังก์ชั่นที่ใช้ในการกำหนดสัญญาณด้านออกที่เรียกว่าฟังก์ชั่นถ่ายโอน (Transfer Function) หรือ ฟังก์ชั่นกระตุ้น (Activation Function)



รูปที่ 2.7 แบบจำลองเซลล์ประสาทเทียม

2.5.6.2 Model Neural Network

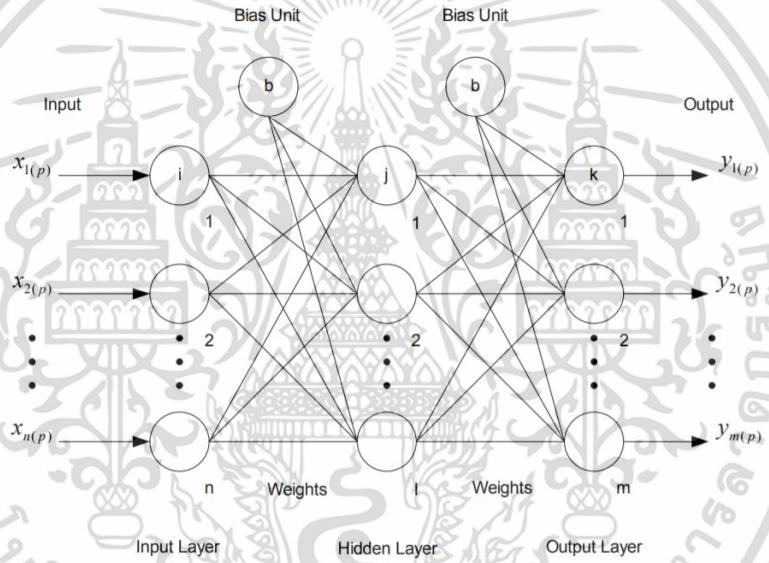
ข่ายงานประสาทแบบป้อนไปหน้า (feed-forward) ประกอบด้วยเซตของบัพ (node) ซึ่งอาจจะถูกกำหนดให้เป็นบัพอินพุต (input nodes) บัพเอาต์พุต (output nodes) หรือ บัพอยู่ระหว่างกลางซึ่งเรียกว่า บัพฮินเดน (hidden nodes) มีการเชื่อมต่อระหว่างบัพ (หรือนิวรอน) โดยกำหนดค่าน้ำหนัก (weight) กำกับอยู่ที่เส้นเชื่อมทุกเส้น เมื่อข่ายงานเริ่มทำงาน จะมีการกำหนดค่าให้แก่บัพอินพุต โดยค่าเหล่านี้ อาจจะได้มาจากการกำหนดโดยมนุษย์ จากเซนเซอร์ที่วัดค่าต่างๆ หรือผลจากโปรแกรมอื่นๆ จากนั้นบัพอินพุต จะส่งค่าที่ได้รับ ไปตามเส้นเชื่อมขาออก โดยที่ค่าที่ส่งออกไปจะถูกคูณกับค่าน้ำหนักของเส้นเชื่อม บัพในชั้นถัดไปจะรับค่า ซึ่งเป็นผลรวมจากบัพต่างๆ แล้วจึงคำนวณผลอย่างง่าย โดยทั่วไปจะใช้ฟังก์ชันซิกมอยด์ (sigmoid function) แล้วส่งค่าไปยังชั้นถัดไป การคำนวณเช่นนี้จะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เกิดขึ้นไปเรื่อยๆ ทีละชั้น จนถึงบัพเอาต์พุต โดยในยุคแรก (ราว ค.ศ. 1970) จำนวนชั้นจะถูกกำหนดไว้เป็นค่าคงที่ แต่ในปัจจุบันมีการนำขั้นตอนวิธีเชิงพันธุกรรม มาช่วยออกแบบโครงสร้างของข่ายงานนิวโรอีโวลูชัน (Neuroevolution)

2.5.6.3 โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (Back Propagation)

โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ ประกอบด้วยชั้นต่างๆ คือชั้นอินพุท (Input layer) กับชั้นเอาท์พุท (Output layer) อย่างละ 1 ชั้น และชั้นซ่อน (Hidden layer) ซึ่งจะอยู่ระหว่างชั้นอินพุทกับชั้นเอาท์พุท สามารถมีกี่ชั้นก็ได้ นอกจากนี้ยังมีหน่วยไบอัส (Bias Unit) ซึ่งจะมีค่าเป็น 1 เสมอ โดยเชื่อมโยงกับทุกๆ โหนดในแต่ละชั้น รูปที่ 2.8 แสดงโครงสร้างของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ



รูปที่ 2.8 โครงสร้างของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (Back Propagation)

2.5.6.4 การเรียนรู้ของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ

การเรียนรู้ของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ อธิบายเป็นขั้นตอนต่างๆ ดังนี้

1. เตรียมข้อมูลที่จะให้โครงข่ายเรียนรู้จากนั้นกำหนดจำนวนโหนดของชั้นอินพุท, จำนวนโหนดของชั้นเอาท์พุท, จำนวนโหนดของชั้นซ่อน, จำนวนชั้นซ่อน, จำนวนรอบสูงสุดในการเรียนรู้และค่าความผิดพลาดที่ยอมรับได้

2. กำหนดค่าโมเมนตัม (Momentum) และอัตราการเรียนรู้ (Learning Rate) การเรียนรู้แบบแพร่ย้อนกลับ จะกำหนดอัตราการเรียนรู้ที่มีค่าน้อยๆ โดยทั่วไปจะอยู่ระหว่าง 0.05 - 0.25 ส่วนค่าโมเมนตัมที่ใช้จะอยู่ในช่วง 0 กับ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. สุ่มค่าถ่วงน้ำหนักให้เส้นเชื่อม ระหว่างโหนดในแต่ละชั้นให้ครบ โดยค่านี้จะเป็นเลขจำนวนจริงที่มีค่าน้อยๆ การกำหนดค่าถ่วงน้ำหนักเริ่มต้นโดยทั่วไปจะสุ่มค่าในช่วงระหว่าง -0.5 และ 0.5 หรือช่วงระหว่าง -1 และ 1 หรืออาจจะอยู่ในช่วงระหว่างค่าใดๆ ที่เหมาะสมตามคุณสมบัติการออกแบบโครงข่ายนั้นๆ

4. รับข้อมูลตัวอย่างเข้าที่โหนดอินพุท เพื่อให้โครงข่ายประสาทเทียมเรียนรู้

5. คำนวณค่าผลลัพธ์ของโครงข่ายประสาทเทียม พร้อมทั้งปรับค่าถ่วงน้ำหนักของโครงข่ายจนกระทั่งโครงข่ายได้ผลลัพธ์ถึงค่าเป้าหมายที่กำหนดไว้ วิธีการคำนวณและสมการที่ใช้ในการหาผลลัพธ์ดังกล่าว มีดังต่อไปนี้

ในส่วนของการคำนวณไปข้างหน้า (Forward Pass) สมการที่ใช้ในการคำนวณหาผลรวมของสัญญาณอินพุทที่เข้าไปยังเซลล์ประสาท j (โหนด j) ในชั้นซ่อนคือสมการที่ 2.16

$$x_j(p) = \sum_{i=1}^{n} x_i(p) w_{ij}(p) + b_j(p) w_j(p) \qquad (2.16)$$

เมื่อ p แทนชุดข้อมูล (Pattern)

n แทนจำนวนโหนดทั้งหมดในชั้นอินพุท

$x_i(p)$ แทนข้อมูลตัวอย่างที่เข้ามาที่โหนดอินพุทตัวที่ i

$w_{ij}(p)$ แทนค่าถ่วงน้ำหนักที่เชื่อมระหว่างโหนดอินพุทตัวที่ i ไปยังโหนด j

$b_j(p)$ แทนไบอัส มีค่าเท่ากับ 1

$w_j(p)$ แทนค่าถ่วงน้ำหนักที่เชื่อมระหว่างไบอัสกับโหนด j

การคำนวณค่าเอาท์พุทของเซลล์ประสาทจะใช้ฟังก์ชันซิกมอยด์ ตามสมการที่ 2.17 มาทำการคำนวณ เนื่องจากสามารถแก้ปัญหาที่มีความซับซ้อนสูงได้ดี กำหนดให้ $y_j(p)$ เป็นเอาท์พุทของชั้นซ่อน จะได้ว่า

$$y_j(p) = \frac{1}{1+e^{-x_j(p)}} \qquad (2.17)$$

เอาท์พุทในชั้นซ่อนจะถูกส่งต่อเป็นอินพุทให้กับชั้นต่อไปจนกระทั่งถึงชั้นสุดท้ายซึ่งก็คือชั้นเอาท์พุท ค่าผลรวมของสัญญาณอินพุทที่เข้าไปยังเซลล์ประสาท k (โหนดk) ที่ชั้นเอาท์พุท $x_k(p)$ และค่าเอาท์พุท $y_k(p)$ คำนวณได้จากสมการที่ 2.18

$$x_k(p) = \sum_{j=1}^{l} y_j(p) w_{jk}(p) + b_k(p) w_k(p) \qquad (2.18)$$

เมื่อ l แทนจำนวนโหนดทั้งหมดในชั้นซ่อน

$y_j(p)$ แทนข้อมูลเอาท์พุทที่ออกจากโหนดในชั้นซ่อนตัวที่ j

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$w_{jk}(p)$ แทนค่าถ่วงน้ำหนักที่เชื่อมระหว่างโหนดชั้นซ่อนตัวที่ j ไปยังโหนด k

$b_k(p)$ แทนไบอัส มีค่าเท่ากับ 1

$w_k(p)$ แทนค่าถ่วงน้ำหนักที่เชื่อระหว่างไบอัส กับ โหนด k

ค่าความคลาดเคลื่อนระหว่างเอาท์พุทที่คำนวณได้กับเอาท์พุทที่ต้องการ $e_k(p)$ ของการฝึกสอนโครงข่ายในแต่ละชุดข้อมูลที่ใช้ฝึกสอน เพื่อนำใช้ในการปรับค่าถ่วงน้ำหนักในชั้นซ่อน สามารถคำนวณได้จากสมการที่ 2.19

$$e_k(p) = y_{d,k}(p) - y_k(p) \tag{2.19}$$

เมื่อ $y_{d,k}(p)$ แทนค่าเอาท์พุทเป้าหมายที่ต้องการ

$y_k(p)$ แทนค่าเอาท์พุทที่คำนวณได้จากโครงข่ายประสาทเทียม

สมการปรับค่าถ่วงน้ำหนักในชั้นเอาท์พุท คำนวณได้จากสมการที่ 2.20

$$w_{jk}(p+1) = w_{jk}(p) + \Delta w_{jk}(p) \tag{2.20}$$

เมื่อ $w_{jk}(p+1)$ แทนค่าถ่วงน้ำหนักของชั้นเอาท์พุทที่ปรับใหม่

$w_{jk}(p)$ แทนค่าถ่วงน้ำหนักเดิมของชั้นเอาท์พุท

$\Delta w_{jk}(p)$ แทนค่าที่ควบคุมปริมาณการปรับค่าถ่วงน้ำหนักในครั้งก่อน (Weight Correction)

ในการปรับสอนจะมีค่าถึงจุดที่ยอมรับได้คือค่าความคลาดเคลื่อนเฉลี่ย (Mean Square Error: MSE) เมื่อค่า MSE ถึงจุดที่ยอมรับ โครงข่ายจะหยุดทำการปรับสอน และนำค่าถ่วงน้ำหนักไปใช้ต่อไป สมการของค่า MSE หาได้จากสมการที่ 2.21

$$MSE = \frac{1}{N}\sum_{k=1}^{N}(y_{d,k(p)} - y_{k(p)})^2 \tag{2.21}$$

2.6 Stacking Ensemble

1.วิธีการแบ็กกิ้ง (Bagging) [11] อาศัยชุดของการฝึกที่แตกต่างกัน ซึ่งถูกสร้างโดยการสุ่มและการแทนที่จากชุดข้อมูลดั้งเดิม สำหรับการฝึก (Bootstraps Replacement) และนำแต่ละชุดของการฝึกทำการฝึกเพื่อให้ได้ตัวจำแนกต่างๆ สำหรับการจำแนกข้อมูลแต่ละตัวจำแนกให้ผลลัพธ์ของแต่ละตัว ต่อมาจะนำผลลัพธ์เหล่านี้ไปผสมกันโดยวิธีการลงคะแนน (vote)

2.วิธีการบูสต์ติง (Boosting) [10] มีหลักการคือจะทำการสร้างโมเดลจำแนกประเภทข้อมูลหลายโมเดล แต่ละโมเดลจะใช้train data ชุดเดียวกันในการสร้าง ซึ่งแต่ละโมเดลจะมีค่าถ่วงน้ำหนัก (Weight)

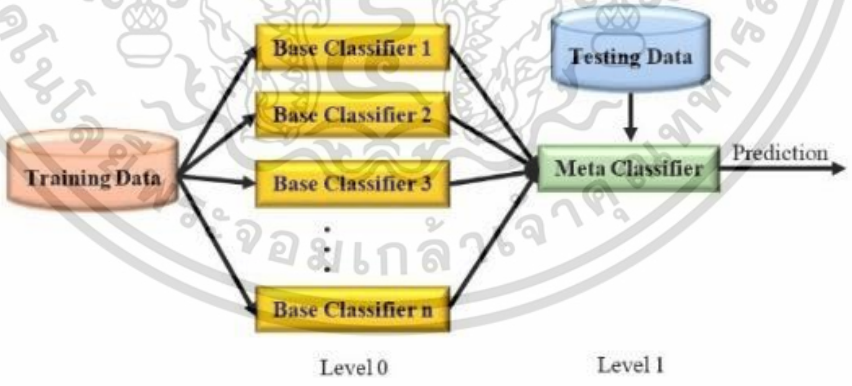
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพิ่มเข้ามา โดยค่าถ่วงน้ำหนักนี้ได้มาจากความถูกต้องของการเรียนรู้บนชุดข้อมูล สำหรับคำตอบสุดท้ายของการทำงานด้วยวิธีบูสต์ติงนั้นจะใช้ วิธีการโหวตแบบถ่วงน้ำหนักแล้วกำหนดกลุ่มให้ข้อมูลใหม่ ด้วยผลโหวตที่มากที่สุด (Majority Voting) วิธีการจำแนกข้อมูลโดยการใช้บูสต์ติงที่มีประสิทธิภาพเป็นที่รู้จักโดยทั่วไปเช่น เอด้าบูสต์ (AdaBoost)

3.วิธีการสแต๊กกิ้ง (Stacking) มีความคล้ายกับ Boosting แต่มีการแบ่งการเรียนรู้เป็นหลายๆ กลุ่ม แล้วเอาข้อมูลทั้งหมดให้กลุ่มแรกเรียน แล้วเอาคำตอบของกลุ่มแรกมารวมกันแล้วส่งต่อให้กลุ่มต่อไปมาเรียนต่อๆกันไป

Stacking เป็นการรวมเทคนิคของตัวแยกประเภทหลายชนิดที่แตกต่างกัน ซึ่งแตกต่างจากวิธีการของ bagging และ boosting ซึ่งวิธีการของ stacking สามารถทำการจำแนก หรือ regression ขึ้นอยู่กับการเลือกตัวแยกประเภท ที่แตกต่างกัน และจำนวนของ stacking folds

เทคนิค Stacking แบ่งเป็นสองชั้นของการจำแนกประเภท ในชั้นแรกจะใช้ตัวแยกประเภทพื้นฐาน (base classifiers) และใช้ตัวแยกประเภทมากกว่าหนึ่งตัว ในชั้นที่สองนั้นได้เรียนรู้ตัวจำแนกจากชั้นแรก (Meta classifier) จากผลของ base classifiers เพื่อให้ได้ผลลัพธ์การจำแนกขั้นสุดท้าย (final classification) ดังนั้นอัลกอริทึม Stacking จะใช้เวลามากกว่าตัวจำแนกอื่น ๆ เช่น bagging และ boosting ในการสร้างแบบจำลอง พูดได้ว่า Stacking คือการนำผลลัพธ์ที่แต่ละโมเดลของ base classifiers (level 0) มาสร้างเป็นอินพุทของชั้น Meta classifiers (level 1) เพื่อสร้างเป็นโมเดลใหม่แล้วนำไปใช้ทำนายกับข้อมูลใหม่ๆ



รูปที่ 2.9 ขั้นตอนการทำงานของ Stacking Ensemble

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
วิธีดำเนินการวิจัย

3.1 ข้อมูลเพลง (Data set)

เพลงที่ใช้ในงานวิจัยนี้เป็นเพลงไทย โดยการโหลดเพลงมาจากอินเตอร์เน็ต โดยทำการจำแนกประเภทของเพลงตามแบบของเพลงไทย ได้แก่ เพลงลูกทุ่ง, เพลงลูกกรุง, เพลงเรกเก้, เพลงสตริงที่มีจังหวะช้า (slow), เพลงที่มีแต่ดนตรี (music box) และเพลงแดนซ์ (EDM) แต่ละประเภทของแนวเพลงใช้อย่างละ 100 เพลง ทั้งหมดมีจำนวน 600 เพลง โดยที่แต่ละเพลงมีความยาว 60 วินาที โดยเลือกข้อมูลในช่วงต้นเพลงมาทำการวิเคราะห์ ในงานวิจัยนี้ใช้ความยาวเพียง 60 วินาทีเนื่องจากใช้เวลาในการประมวลผลไม่นานมากนัก

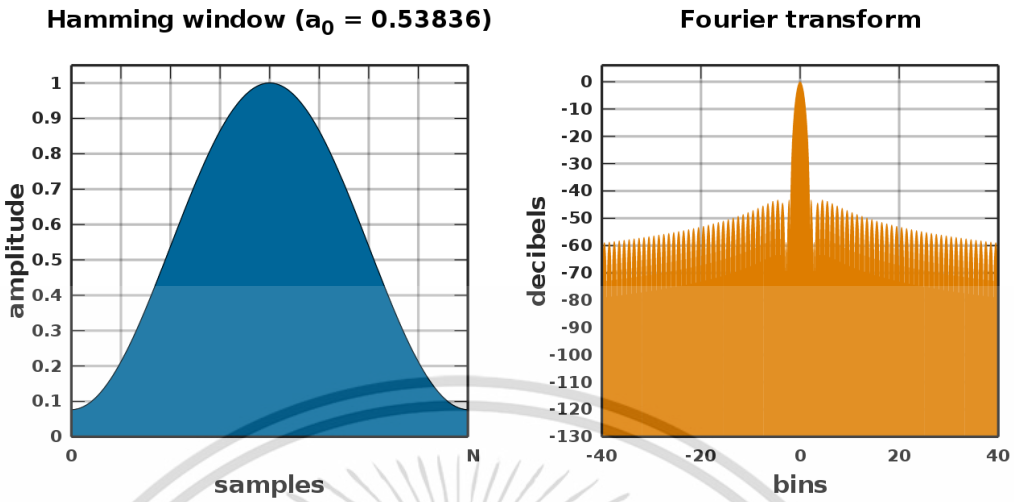
3.2 การดึงคุณลักษณะ (Feature Extraction)

รูปที่ 3.1 แผนภาพการแยกคุณลักษณะของข้อมูล

ในขั้นตอนของการดึงฟีเจอร์ ทำการแบ่งสัญญาณด้วยเทคนิคหน้าต่าง (Hamming Window) เพื่อทำให้เกิดสัญญาณย่อยเป็นช่วงที่เท่าๆกัน จากนั้นนำสัญญาณภายใน หน้าต่างมาผ่านการแปลงฟูริเยร์แบบเร็ว (Fast Fourier Transform : FFT) เพื่อให้ได้ข้อมูลของเสียงในโดเมนความถี่ และนำข้อมูลความถี่ของเสียงมาหาฟีเจอร์ต่างๆ ได้แก่ Spectral Centroid, Spectral Flux, Spectral RollOff, Spectral Contrast, Spectral Mean

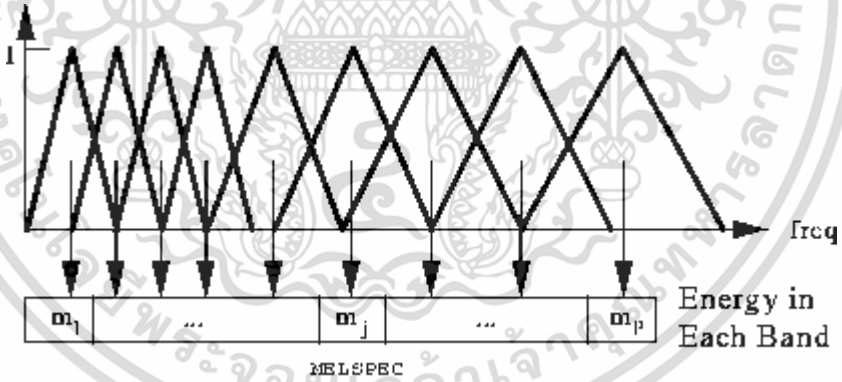
ในการหาฟีเจอร์ MFCC ทำโดยนำข้อมูลความถี่ของเสียงมาผ่านการกรองด้วยชุดตัวกรองแบบสามเหลี่ยมที่สร้างจากสเกลเมล (Mel scale) ซึ่งการใช้ตัวกรองแบบสามเหลี่ยมที่สร้างจากสเกลเมลเป็นเทคนิคที่ใช้ปรับสเกลของสเปคตรัมให้อยู่บนสเกลที่เหมาะสม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.2 ด้านซ้ายคือกราฟหน้าต่าง Hamming Window และ ด้านขวาคือ Fourier Transform ของ Hamming Window

ในการหาฟีเจอร์ MFCC ทำโดยนำข้อมูลความถี่ของเสียงมาผ่านการกรองด้วยชุดตัวกรองแบบสามเหลี่ยมที่สร้างจากสเกลเมล (Mel scale) ซึ่งการใช้ตัวกรองแบบสามเหลี่ยมที่สร้างจากสเกลเมลเป็นเทคนิคที่ใช้ปรับสเกลของสเปคตรัมให้อยู่บนสเกลที่เหมาะสม

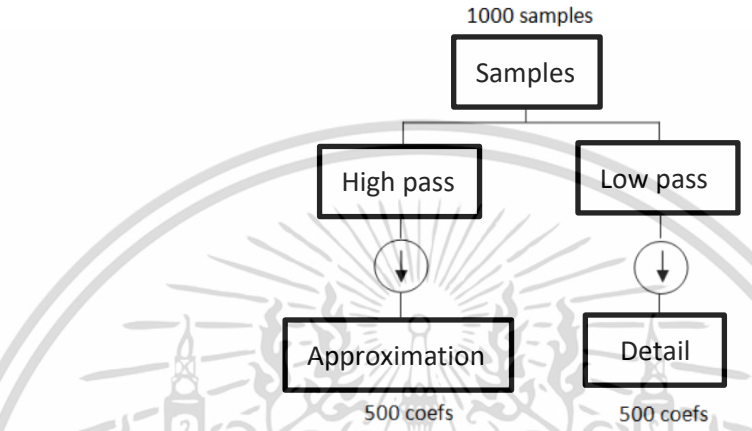


รูปที่ 3.3 Mel Scale Filter Bank

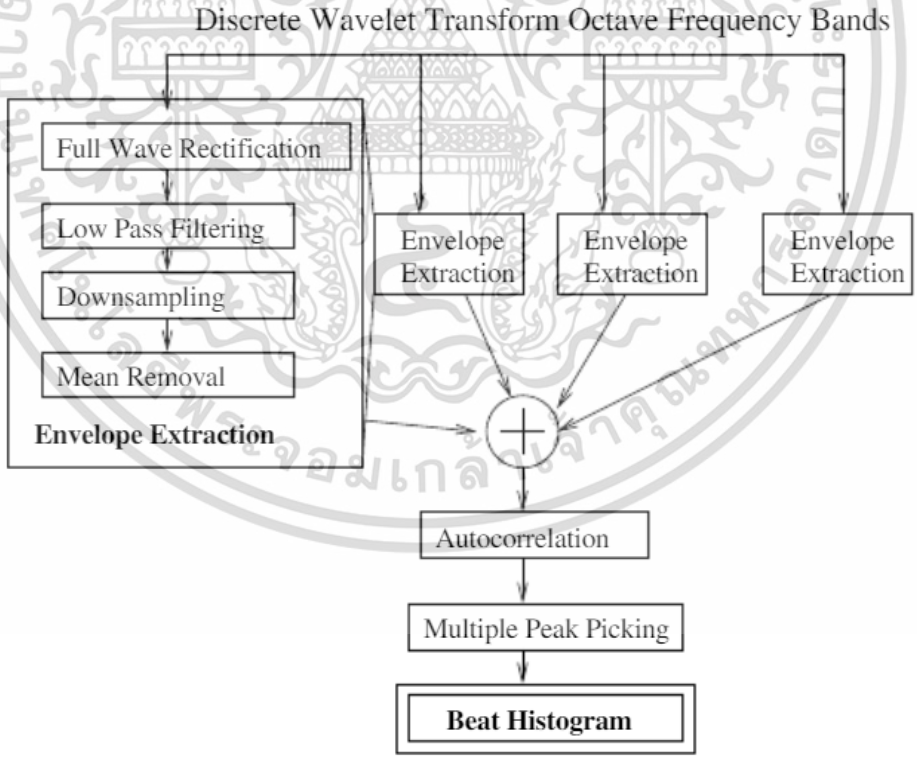
ในการหา Beat Histogram การคำนวณฟีเจอร์นี้รวมถึงการตรวจจับจังหวะ โดยแยกสัญญาณออกเป็นความถี่แบบ Octave Bands แล้วใช้การแปลงสัญญาณเวฟเลตแบบเต็มหน่วย โดยนำสัญญาณมาผ่านตัวกรองความถี่ต่ำ (low pass filter) และความถี่สูง (high pass filter) เพื่อทำการแยกองค์ประกอบสัญญาณออกเป็นส่วนที่มีความถี่ต่ำ (approximation) และส่วนที่มีความถี่สูง (detail) โดยสัญญาณที่ได้มานั้นจะผ่านอัตราสุ่มลง (down sampling) ด้วย 2 ทำให้จำนวนตัวอย่าง (sample) ของผลลัพธ์ที่ได้ลดลงครึ่งหนึ่งของสัญญาณเดิม ดังแสดงในรูปที่ 3.4 หลังจากนั้นนำสัญญาณที่ผ่านการทำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กระบวนการที่กล่าวมาแล้วมาลบค่าเฉลี่ยของสัญญาณของแต่ละย่านแล้วทำการสัญญาณที่ทำการ Envelope ของแต่ละย่านแล้วเข้าด้วยกัน จากนั้นคำนวณค่าความสัมพันธ์ของสัญญาณด้วยวิธี autocorrelation แล้วทำการเลือกค่าที่สูงสุดของแอมพลิจูดที่โดดเด่นของ autocorrelation นำมาเรียงเป็นฮิสโตแกรมจะได้ Beat Histogram ตามรูปที่ 3.5



รูปที่ 3.4 การแปลงเวฟเลตแบบเต็มหน่วย (Discrete wavelet transform)



รูปที่ 3.5 แผนภาพแสดงการคำนวณหา Beat Histogram

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 ฟีเจอร์ต่างๆ ที่ถูกนำมาใช้แยกคุณลักษณะของข้อมูล

Feature	Description
1	Mean of Spectral Centroid
2	Standard of Spectral Centroid
3	Mean of Spectral Flux
4	Standard of Spectral Flux
5	Mean of Spectral Rolloff
6	Standard of Spectral Rolloff
7	Mean of Volume
8	Standard of Volume
9	Mean of Zero Crossing
10	Standard of Zero Crossing
11	Mean of Short Time Energy
12	Standard of Short Time Energy
13	Low Energy
14	Mean of first Mel-Frequency Cepstral Coefficients
15	Mean of second Mel-Frequency Cepstral Coefficients
16	Mean of third Mel-Frequency Cepstral Coefficients
17	Mean of fourth Mel-Frequency Cepstral Coefficients
18	Mean of fifth Mel-Frequency Cepstral Coefficients
19	Standard of first Mel-Frequency Cepstral Coefficients
20	Standard of second Mel-Frequency Cepstral Coefficients

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 (ต่อ) ฟีเจอร์ต่างๆ ที่ถูกนำมาใช้แยกคุณลักษณะของข้อมูล

Feature	Description
21	Standard of third Mel-Frequency Cepstral Coefficients
22	Standard of fourth Mel-Frequency Cepstral Coefficients
23	Standard of fifth Mel-Frequency Cepstral Coefficients
24	Period Peak Unfolded of Pitch
25	Period Peak Folded of Pitch
26	Amplitude Maximum Folded of Pitch
27	Strength of Pitch
28	Beat Per Minute (BPM)
29	First Histogram Peak of Beat
30	Second Histogram Peak of Beat
31	Ratio Amplitude Histogram of Beat
32	Strength of Beat
33	Mean of Spectral Contrast
34	Standard of Spectral Contrast
35	Mean of Spectral Mean
36	Standard of Spectral Mean
37	Mean of Frequency Histogram
38	Standard of Frequency Histogram
39	Mean of Pitch
40	Standard of Pitch
41	Mean of Pitch Strength
42	Standard of Pitch Strength
43	Sum of Pitch Strength

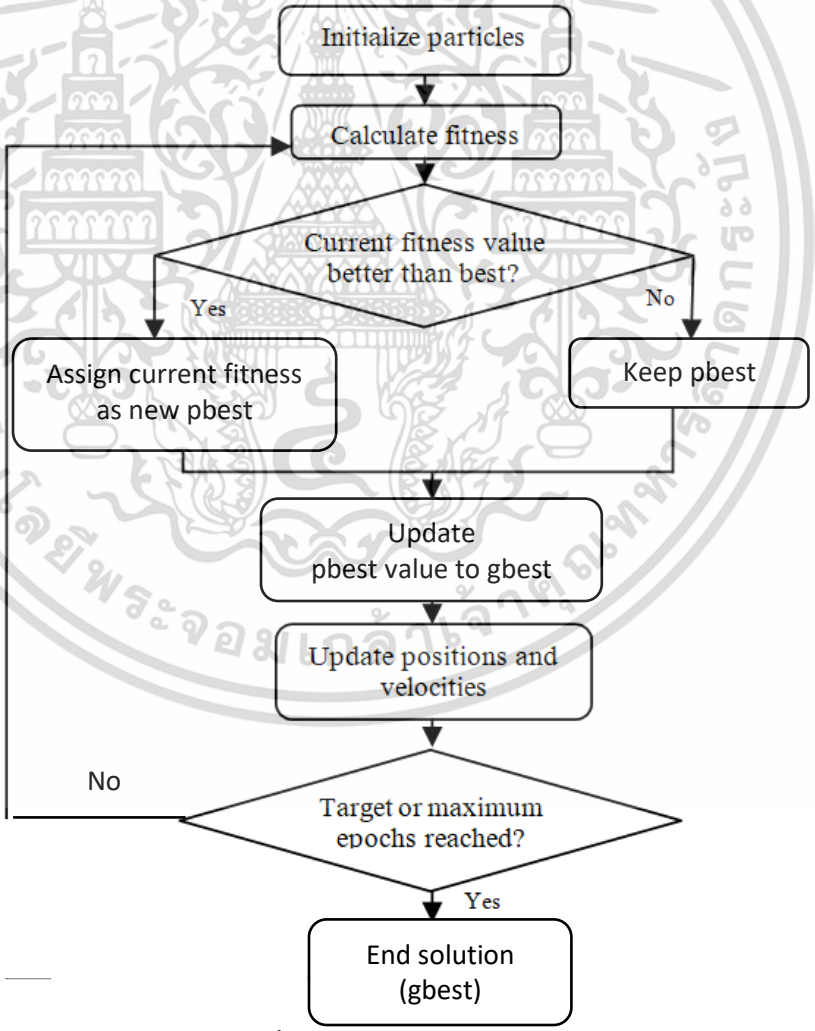
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 วิธีการคัดเลือกคุณลักษณะ (Feature selection)

ในการคัดเลือกคุณลักษณะที่โดดเด่นมากจากคุณลักษณะทั้งหมด ในงานวิจัยนี้ใช้วิธีการของ Particle Swarm Optimization (PSO) เนื่องจากง่ายต่อการใช้งานและทำงานอย่างมีประสิทธิภาพสำหรับการแก้ปัญหาอย่างต่อเนื่อง แต่วิธีนี้มีข้อเสียคือค่าจะลู่เข้าก่อนเวลาที่สมควร

วิธีการของ PSO เป็นแบบ wrapper approach คือเป็นการคัดเลือกฟีเจอร์ด้วยการสร้างโมเดล (classification model) ฟีเจอร์ขึ้นมาจากเซตของฟีเจอร์ที่กำหนดไว้และวัดประสิทธิภาพการทำงานของโมเดล และเลือกเซตของที่ทำให้โมเดลมีประสิทธิภาพมากที่สุดมาใช้งาน เช่น โมเดลที่ให้ค่าความถูกต้อง (accuracy) มากที่สุด โดยในงานวิจัยนี้ใช้ข้อมูลเทรน (Train data) 480 และข้อมูลทดสอบ (Test data) 120 โมเดลของ PSO ที่ใช้เป็นแบบ continuous

รูปข้างล่างแสดงแผนภาพลำดับขั้นตอนการทำงานของ PSO ลำดับการทำงานเป็นดังนี้



รูปที่ 3.6 ขั้นตอนกระบวนการทำ PSO

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.ตั้งค่าเริ่มต้นของตัวแปรต่างๆ ได้แก่ ค่าคงที่ (C_1, C_2) เป็น 2, ค่าถ่วงน้ำหนัก (w) เป็น 0.2, จำนวนรอบ 100 รอบ, จำนวนฟีเจอร์ (3, 5, 7, 9), จำนวนประชากร (20, 50, 80, 100)

2.คำนวณค่า fitness จากโมเดลที่กำหนดไว้ ได้แก่ kNN, SVM และ DT โดยทำทีละโมเดล ซึ่ง fitness value คำนวณได้จากค่าความถูกต้องของแต่ละโมเดล

3.พิจารณาว่าค่า fitness ดีกว่าค่า fitness เดิมหรือไม่ ถ้าค่า fitness ค่าใหม่ดีกว่าค่าเดิม ให้เก็บเป็นค่าใหม่ของ pbest ถ้าไม่ ให้เก็บ pbest ค่าเดิม

4.ทำจนครบจำนวนประชากรที่กำหนดไว้ โดยที่ประชากรแต่ละตัวจะมีจำนวนฟีเจอร์ที่ถูกเลือกเท่ากัน

5.ทำการอัพเดต pbest ค่าที่ดีที่สุดเป็น gbest

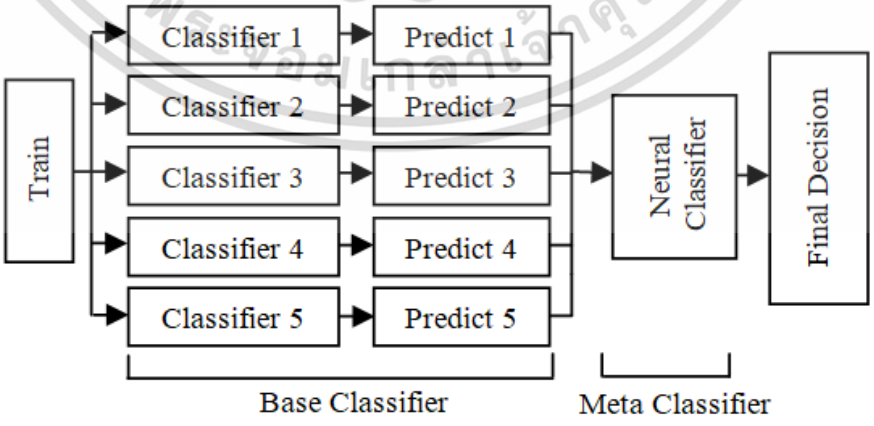
6.เมื่อจบการทำงานจะได้ค่า gbest ที่ดีที่สุดของรอบที่ 1 โดยคัดเลือกจากจำนวนประชากรที่ตั้งไว้

7.อัพเดตตำแหน่งและความเร็วตามสมการที่ 2.8 และ 2.9 ตามลำดับ โดยครั้งแรกจะทำการสุ่มฟีเจอร์มาก่อน หลังจากทำจนได้ค่า gbest จึงอัพเดตตำแหน่งและความเร็วใหม่ โดยที่ฟีเจอร์ที่ดีก็คงไว้ ส่วนฟีเจอร์ที่แย่ ก็สุ่มฟีเจอร์ใหม่มาแทนที่

8.ทำซ้ำเดิมจนครบจำนวนรอบที่กำหนดไว้ (รอบถัดไปถ้า gbest มีค่าดีกว่าค่า gbest เดิม ให้อัพเดตค่าเป็น gbest ใหม่) จะได้ค่า gbest ที่ดีที่สุดของแต่ละโมเดล

3.4 Stacking Ensemble

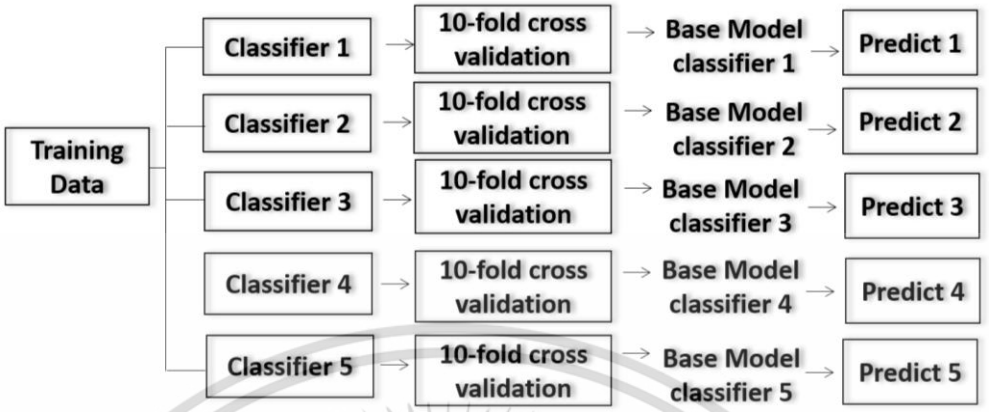
หลังจากทำ PSO แล้วจะได้ฟีเจอร์ที่ถูกเลือกมา จากนั้นนำมาเข้ากระบวนการของ stacking ensemble รูปข้างล่างแสดงแผนภาพลำดับขั้นตอนการทำงานของ Stacking Ensemble ลำดับการทำงานเป็นดังนี้



รูปที่ 3.7 แผนภาพแสดงวิธีการของ Stacking Ensemble

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.1 Base Classifier



รูปที่ 3.8 วิธีการทำงานของ Base Classifier

1.นำข้อมูลที่ผ่านขั้นตอนการทำ PSO มาแบ่งเป็น Train data 80% และ Test data 20%

2. ในส่วนของข้อมูลที่นำมาเทรนนำเข้าสู่วิธีการของ k-fold cross validation โดยแบ่งข้อมูลออกเป็น 2 ส่วน คือ ส่วนที่เทรน 72 และส่วนที่ทดสอบ 8

3.นำข้อมูลเทรนมาใส่ Classified แต่ละตัว ซึ่งตัว Classifier ที่ใช้ในงานวิจัยนี้มีด้วยกัน 5 ตัวได้แก่ Random Forest, kNN, SVM, DT และ Naive Bayes

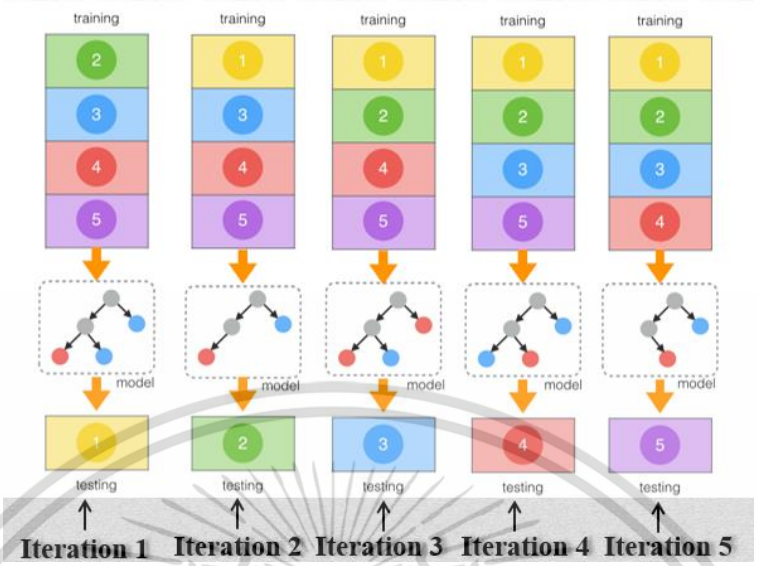
4.จากนั้นนำโมเดลที่ได้จากการเทรนข้อมูลของ Classified แต่ละตัว มาทำนายผลลัพธ์

5.นำผลลัพธ์ที่ได้จากการทำนายเข้าสู่กระบวนการ Meta Classifier

3.4.2 วิธีการของ k-fold cross validation

Cross Validation เป็นวิธีที่นิยมในการทำงานวิจัย เพื่อใช้ในการทดสอบประสิทธิภาพของโมเดล เนื่องจากผลที่ได้มีความน่าเชื่อถือ การวัดประสิทธิภาพด้วยวิธี Cross-validation จะทำการแบ่งข้อมูลออกเป็นหลายส่วน (มักจะแสดงด้วยค่า k) เช่น 5-fold cross validation คือ ทำการแบ่งข้อมูลออกเป็น 5 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดล ทำวนไปเช่นนี้จนครบจำนวนที่แบ่งไว้ ในงานวิจัยนี้ใช้ 10-fold cross validation

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.9 วิธีการทำงานของ k-fold cross validation

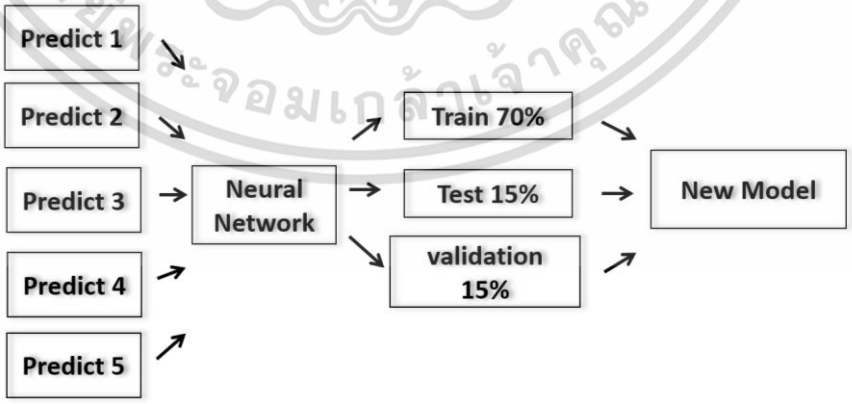
จากรูปที่ 3.9 แสดงวิธีการทำงานของ k-fold cross validation

1.แบ่งข้อมูลออกเป็น 10 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่าๆ กัน

2.แบ่งข้อมูลไว้หนึ่งชุดสำหรับเป็นข้อมูลทดสอบ ข้อมูลอีก 9 ชุดที่เหลือเป็นข้อมูลเทรน เช่น ข้อมูลชุดที่ 2, 3, 4, 5, ,6, 7, 8, 9, 10 เป็นข้อมูลเทรน ข้อมูลชุดที่ 1 เป็นข้อมูลทดสอบ

3.ทำซ้ำขั้นตอนที่ 2 แต่เปลี่ยนชุดข้อมูลทดสอบ เช่น ชุดที่ 1, 3, 4, 5, 6, 7, 8, 9, 10 เป็นข้อมูลเทรน ข้อมูลชุดที่ 2 เป็นข้อมูลทดสอบ ทำซ้ำจนกระทั่งครบจำนวน 10 รอบ

3.4.3 Meta Classifier



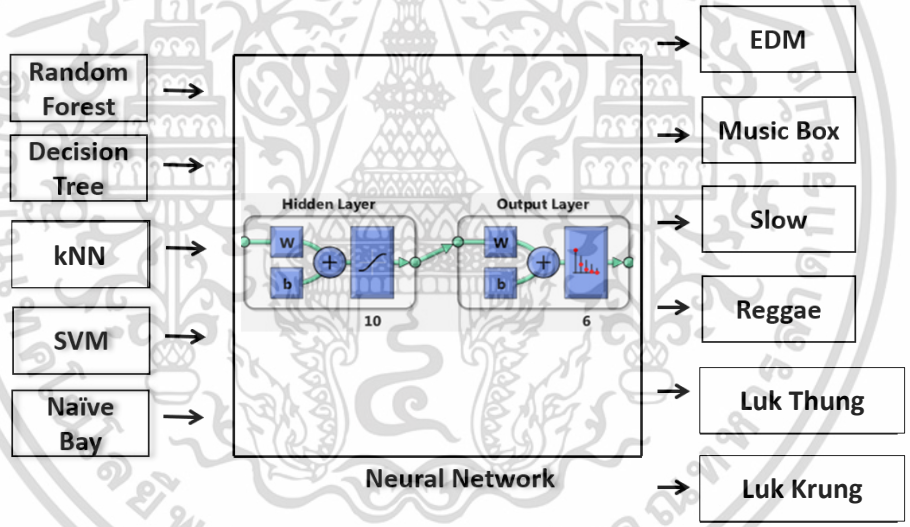
รูปที่ 3.10 วิธีการทำงานของ Meta Classifier

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นำผลลัพธ์ที่ได้จากการทำนายของกระบวนการ Base Classifier เป็นข้อมูลเทรนชุดใหม่ของขั้นตอนการทำ Meta Classifier แล้วนำเข้า Neural Network โดยใน Neural มีการแบ่งข้อมูลเป็นข้อมูลเทรน 70%, ข้อมูลทดสอบ 15%, ข้อมูลตรวจสอบ (Validation) 15% เมื่อสิ้นสุดกระบวนการจะได้โมเดลสำหรับการทำนายข้อมูลทดสอบ

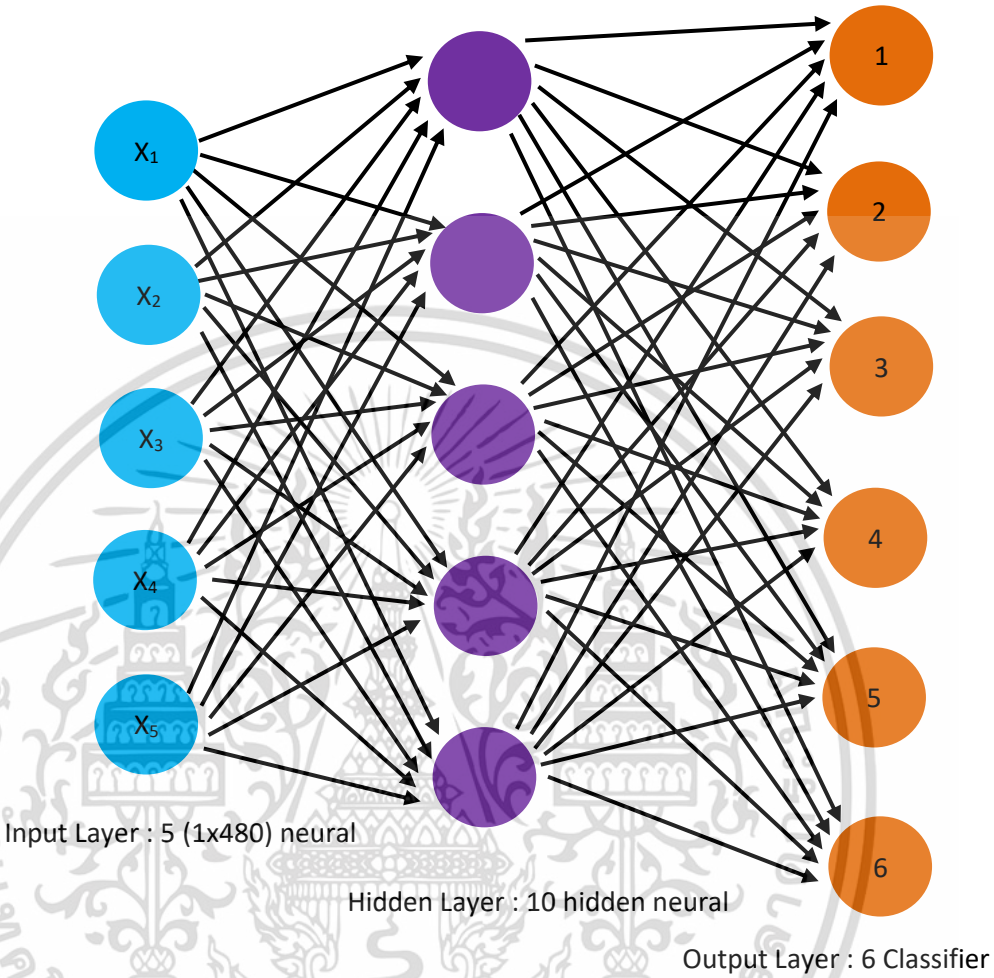
3.5 Neural Network

ขั้นตอนการทำนิวรอน นำผลลัพธ์ที่ได้จากการทำนายในกระบวนการทำ Base Classifier มาตั้งต้นเป็นอินพุทของนิวรอน ซึ่งผลลัพธ์ที่ได้จากการทำนายมาจากตัวจำแนกประเภทแบบ Random Forest, kNN, SVM, DT และ Naive Bayes โดยมี label เดิมเป็นตัวกำหนดว่าแต่ละการทำนายควรจัดอยู่ในกลุ่มไหน ดังรูปที่ 3.11 มีฟังก์ชันทรานเฟอร์คือ ซิกมอยด์ (Sigmoid) ใช้นิวรอนแบบ pattern recognition [6]



รูปที่ 3.11 วิธีการทำงานของ Neural Network

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.12 โครงข่ายของ Neural Network

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
การทดลองและผลการทดลอง

4.1 Feature Extraction

ในขั้นตอนการแยกฟีเจอร์เราแบ่งชุดข้อมูลโดยการใช้ Hamming Window ซึ่งมีหน้าต่างละ 0.5 วินาที หน้าต่างถัดไปจะมีการเลื่อนทับหน้าต่างเดิม 0.25 วินาที ในการแยกฟีเจอร์สามารถแบ่งฟีเจอร์ได้เป็น 3 กลุ่มใหญ่ๆ คือ Beat, Timbral texture และ pitch เมื่อทำการแยกฟีเจอร์ออกมาจะได้ผลลัพธ์เป็นฟีเจอร์จำนวน 43 ฟีเจอร์

ผลลัพธ์ที่ได้จากการแยกลักษณะฟีเจอร์ของ Beat จะได้ ฟีเจอร์ออกมาจำนวน 5 พารามิเตอร์ด้วยกัน คือ beat per minute (bmp), first histogram peak, second histogram peak, ratio amplitude histogram, and beat strength [1]

ผลลัพธ์ที่ได้จากการแยกลักษณะฟีเจอร์ของ Timbral จะได้ ฟีเจอร์ออกมาจำนวน 29 พารามิเตอร์ ประกอบไปด้วย low energy, mean and standard deviation (std) of spectral centroid, spectral flux, spectral roll off, spectral contrast, spectral mean, frequency histogram [1 4], volume, zero-crossing, energy, first mel-frequency cepstral coefficients (mfcc), second mfcc. Third mfcc, fourth mfcc, fifth mfcc [1].

ผลลัพธ์ที่ได้จากการแยกลักษณะฟีเจอร์ของ Pitch จะได้ ฟีเจอร์ออกมาจำนวน 9 พารามิเตอร์ ประกอบไปด้วย all strength of pitch, sum of pitch strength, period peak unfolded, period peak unfolded, amplitude maximum folded [1], mean, standard of pitch และ pitch strength.

กราฟข้างล่างแสดงค่าฟีเจอร์แต่ละชนิดที่ได้จากการทำ feature extraction โดยมีกราฟแนวตั้งเป็นค่าแอมพลิจูดของแต่ละเพลง ส่วนกราฟแนวนอนแสดงถึงตำแหน่งของเพลง

เพลงที่ 1-100 เป็นประเภทแนวเพลงแบบ EDM

เพลงที่ 101-200 เป็นประเภทแนวเพลงแบบ Music Box

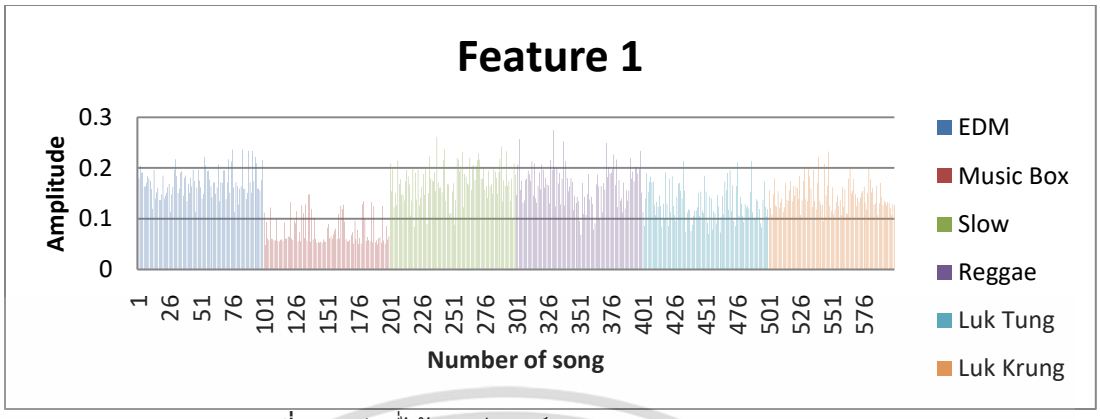
เพลงที่ 201-300 เป็นประเภทแนวเพลงแบบ Slow

เพลงที่ 301-400 เป็นประเภทแนวเพลงแบบ Reggae

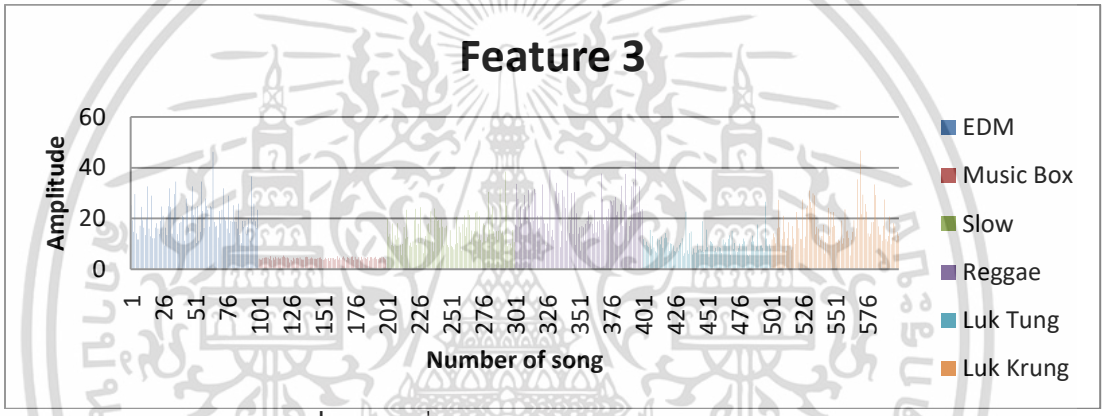
เพลงที่ 401-500 เป็นประเภทแนวเพลงแบบ Luk Tung

เพลงที่ 501-600 เป็นประเภทแนวเพลงแบบ Luk Krung

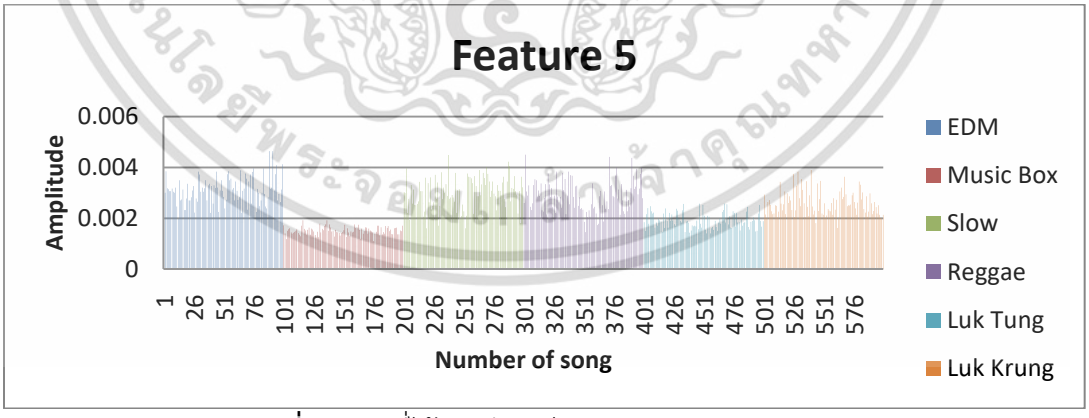
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.1 ค่าที่ได้จากฟีเจอร์ Mean Spectral Centroid

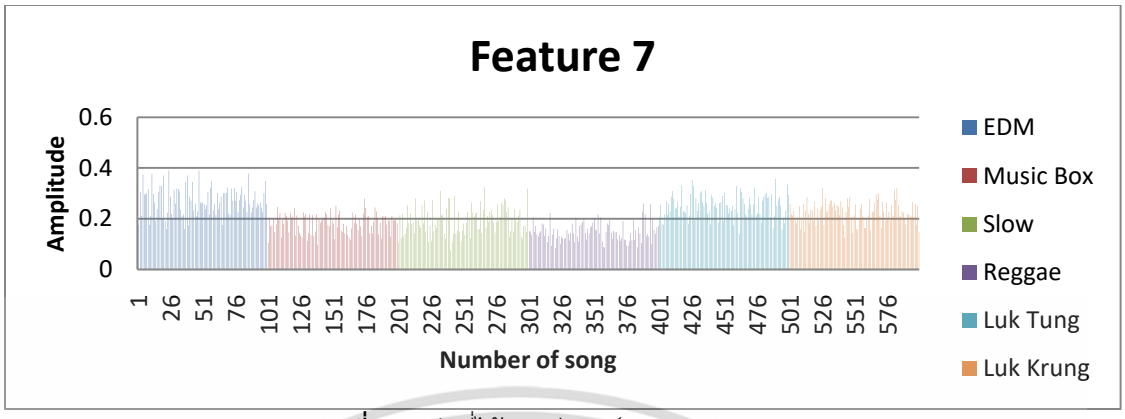


รูปที่ 4.2 ค่าที่ได้จากฟีเจอร์ Mean Spectral Flux

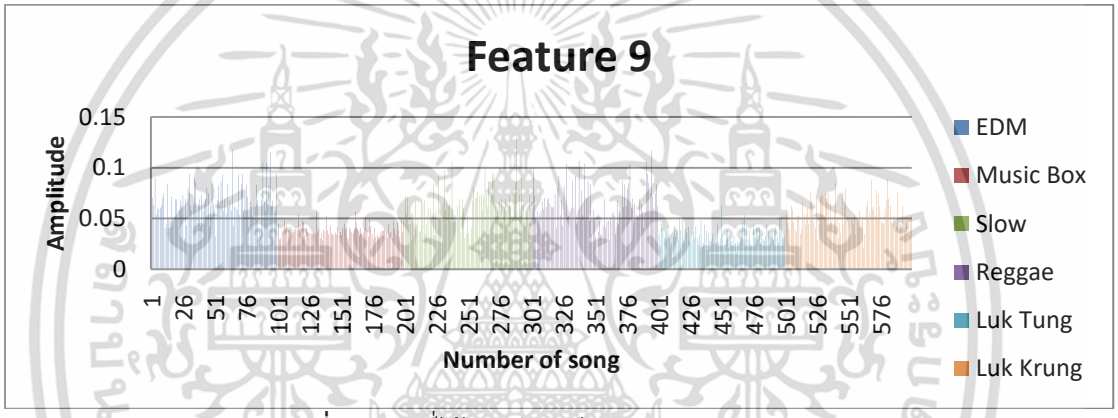


รูปที่ 4.3 ค่าที่ได้จากฟีเจอร์ Mean Spectral Rolloff

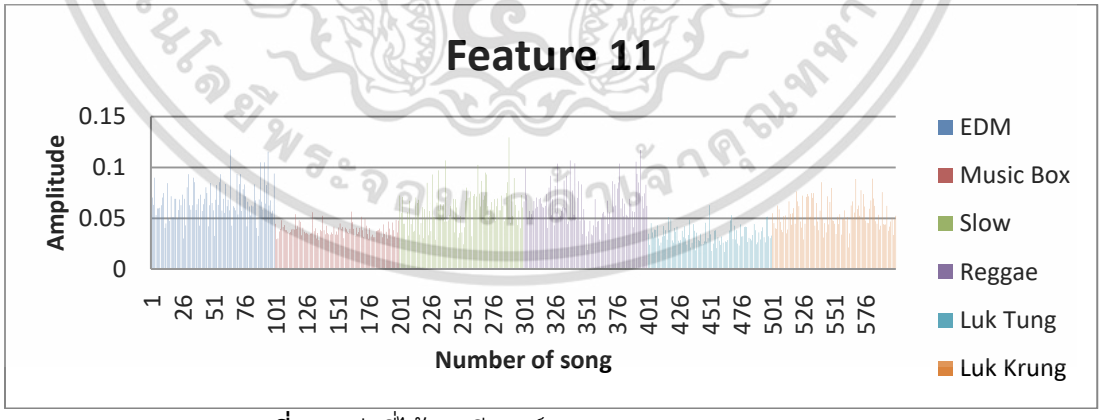
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.4 ค่าที่ได้จากฟีเจอร์ Mean Volume

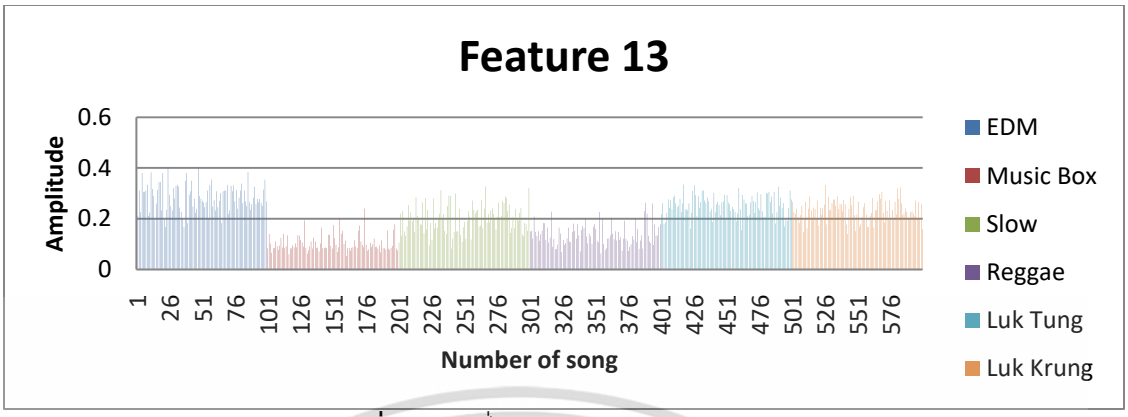


รูปที่ 4.5 ค่าที่ได้จากฟีเจอร์ Mean Zero Crossing

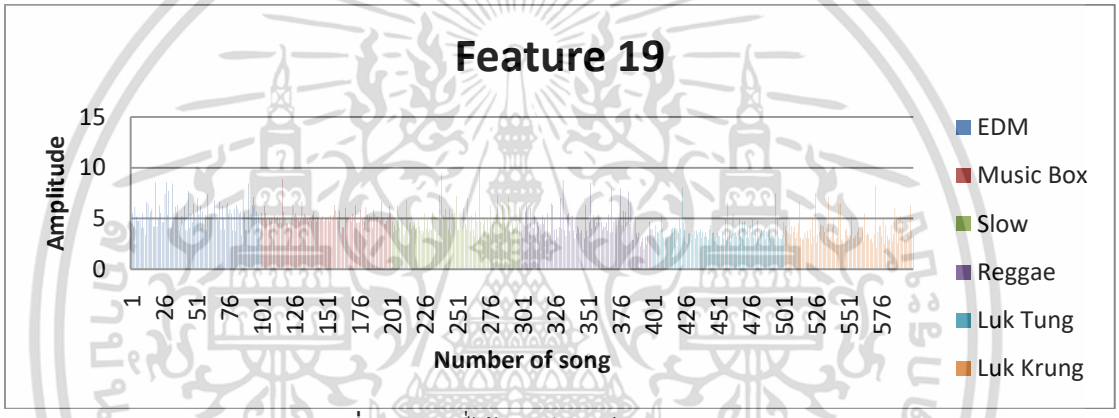


รูปที่ 4.6 ค่าที่ได้จากฟีเจอร์ Mean Short Time Energy

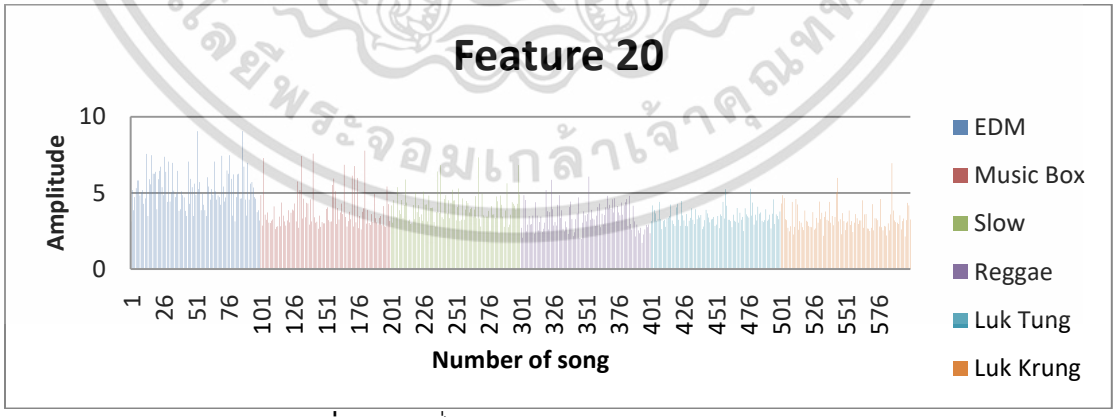
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.7 ค่าที่ได้จากฟีเจอร์ Low Energy

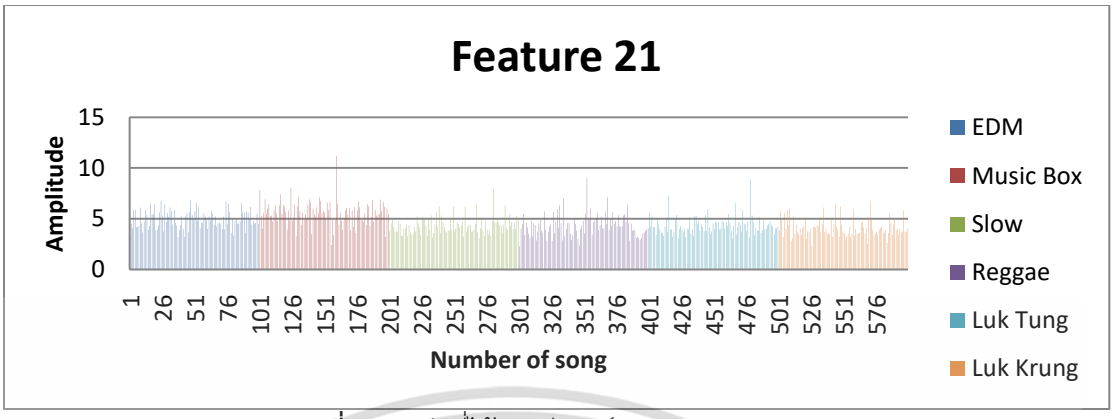


รูปที่ 4.8 ค่าที่ได้จากฟีเจอร์ Std First MFCC

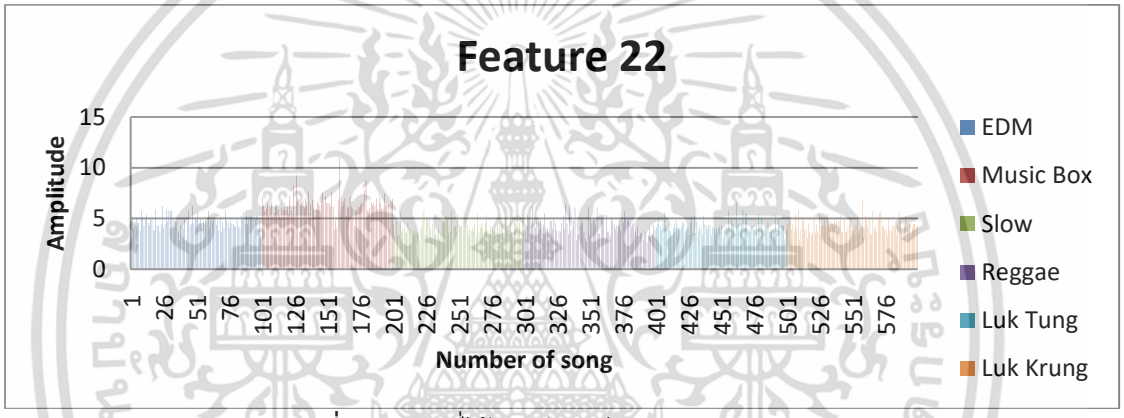


รูปที่ 4.9 ค่าที่ได้จากฟีเจอร์ Std Second MFCC

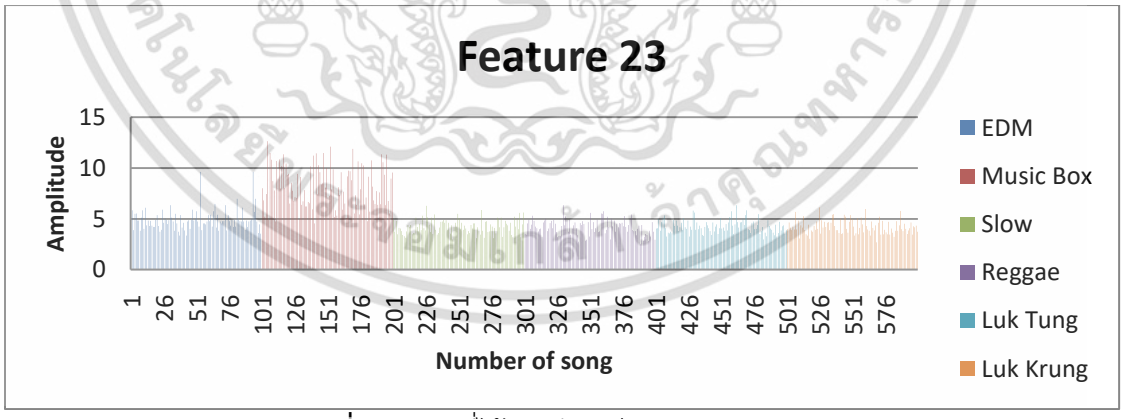
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.10 ค่าที่ได้จากฟีเจอร์ Std Third MFCC

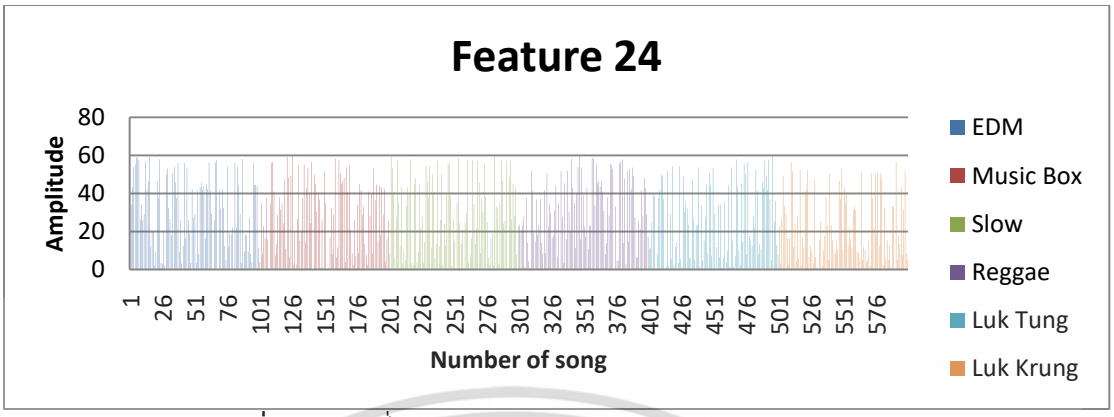


รูปที่ 4.11 ค่าที่ได้จากฟีเจอร์ Std Fourth MFCC

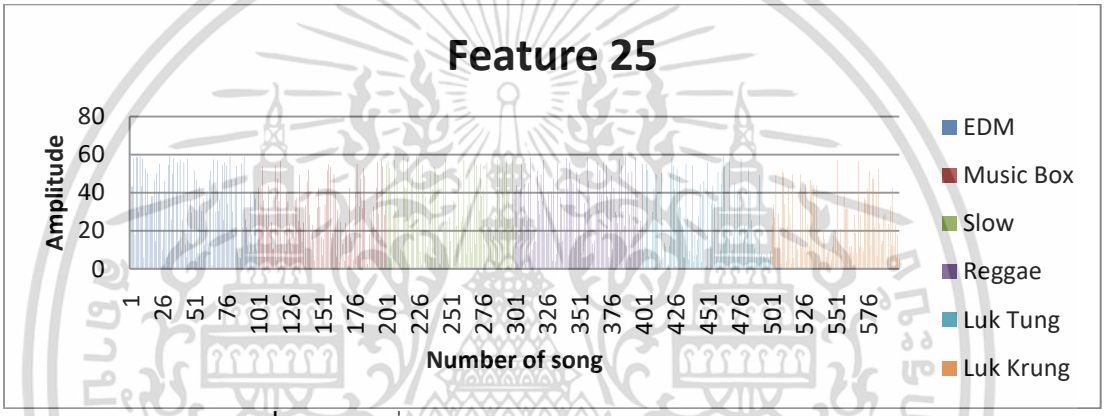


รูปที่ 4.12 ค่าที่ได้จากฟีเจอร์ Std Fifth MFCC

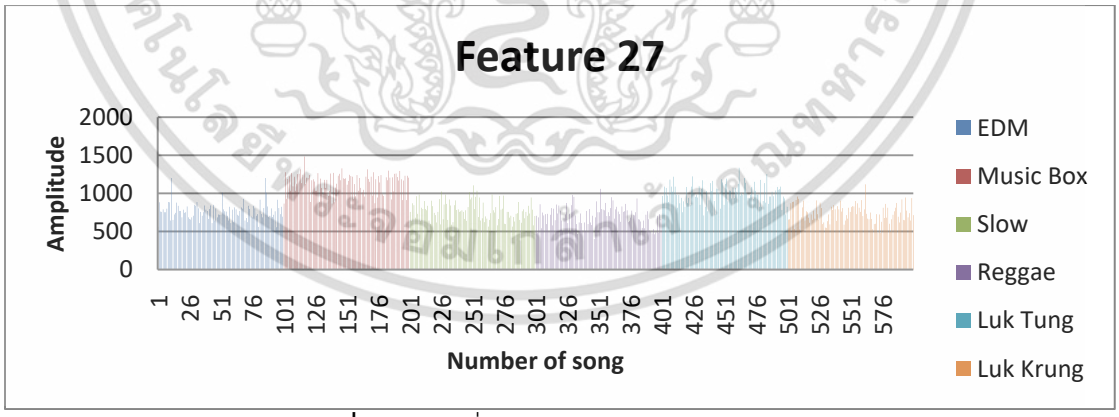
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.13 ค่าที่ได้จากฟีเจอร์ Period Peak Unfolded Pitch

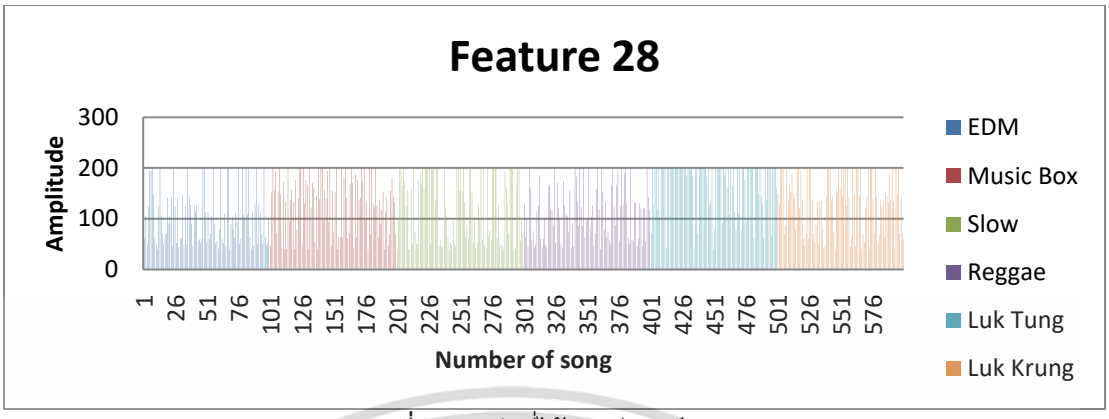


รูปที่ 4.14 ค่าที่ได้จากฟีเจอร์ Period Peak Folded Pitch

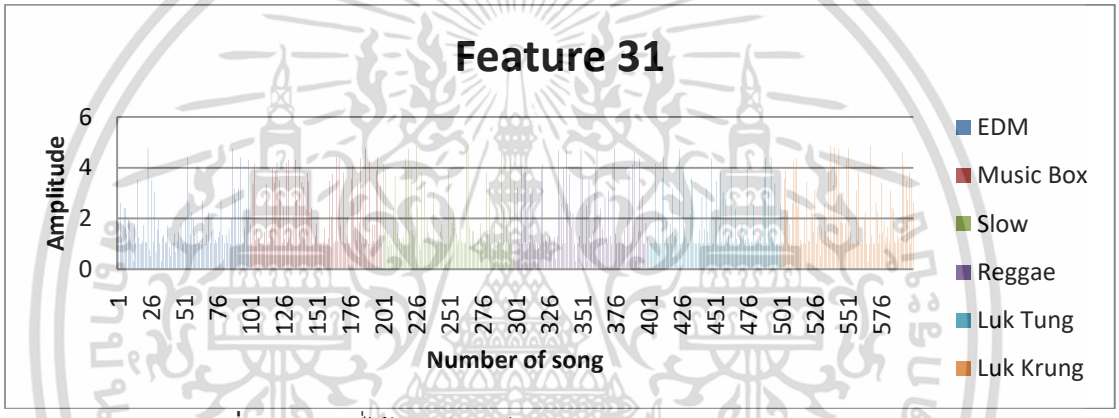


รูปที่ 4.15 ค่าที่ได้จากฟีเจอร์ Strength Pitch

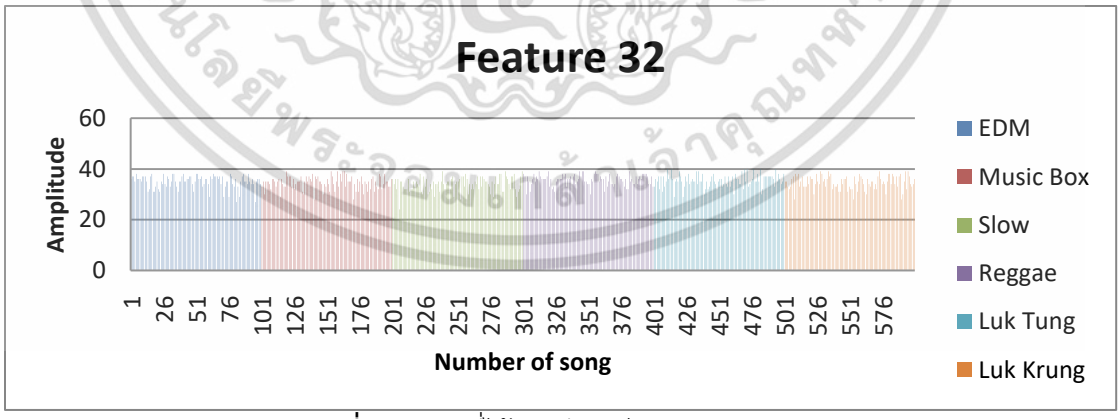
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.16 ค่าที่ได้จากฟีเจอร์ BPM

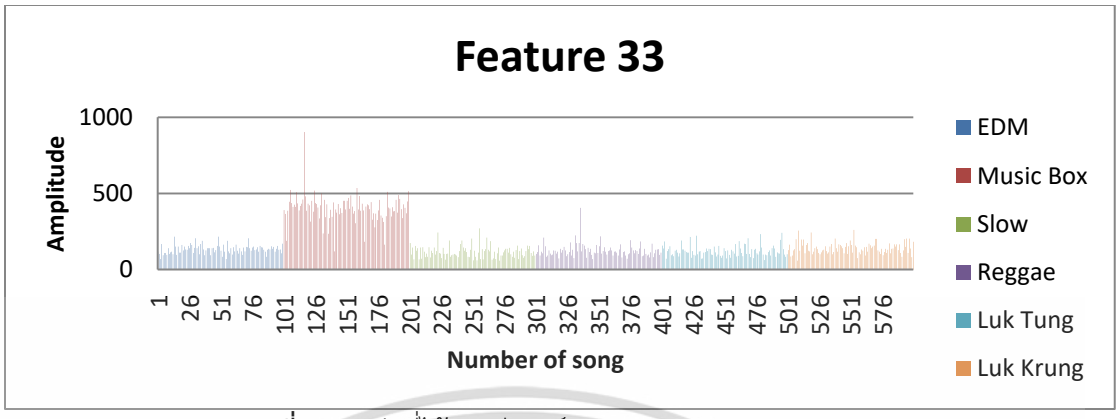


รูปที่ 4.17 ค่าที่ได้จากฟีเจอร์ Ratio Amplitude Histogram Beat

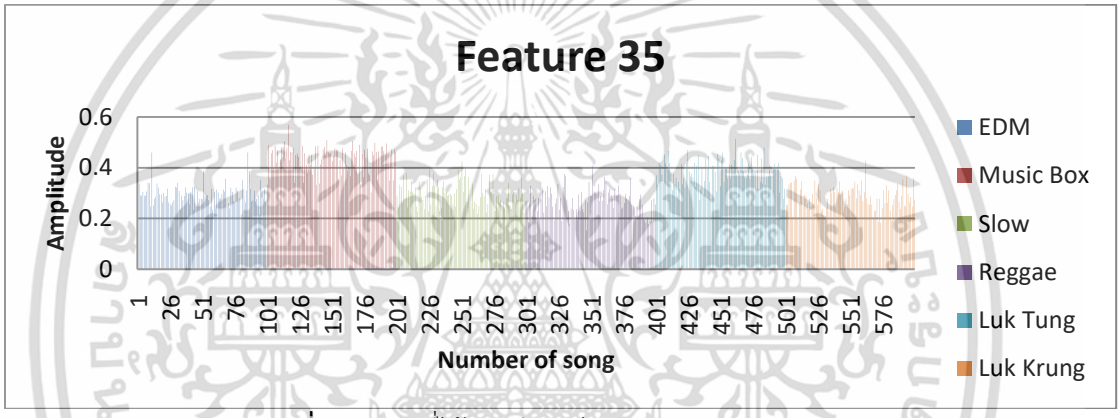


รูปที่ 4.18 ค่าที่ได้จากฟีเจอร์ Beat Strength

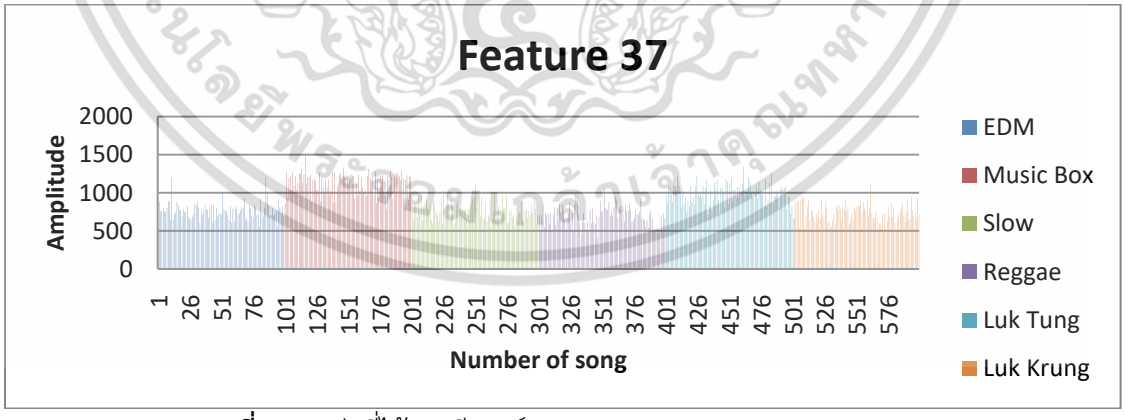
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.19 ค่าที่ได้จากฟีเจอร์ Mean Spectral Contrast

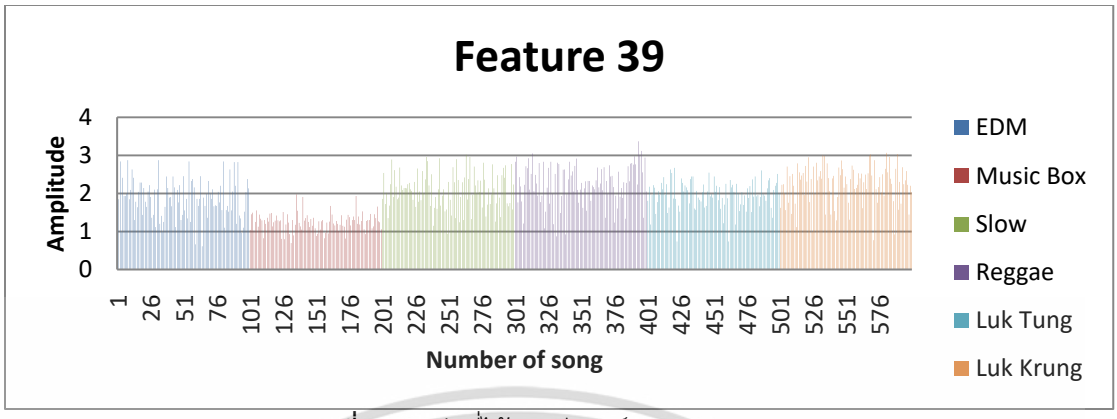


รูปที่ 4.20 ค่าที่ได้จากฟีเจอร์ Mean Spectral Mean

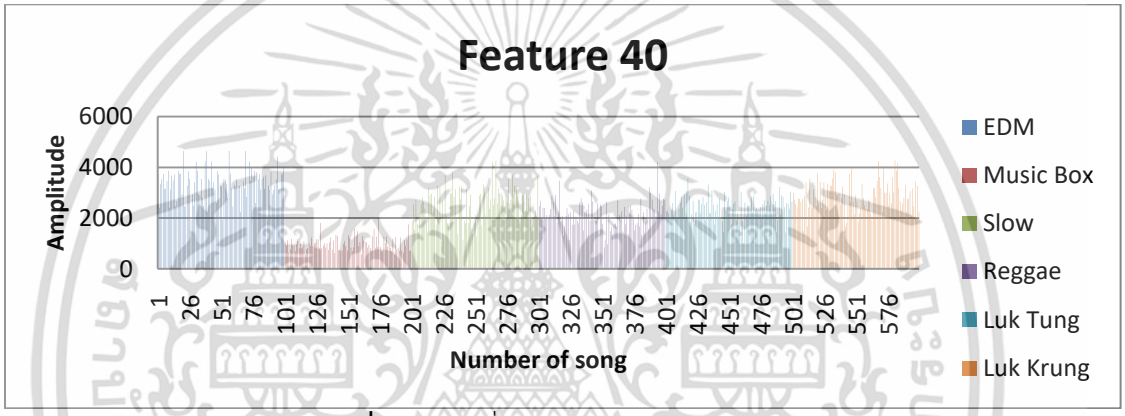


รูปที่ 4.21 ค่าที่ได้จากฟีเจอร์ Mean of Frequency Histogram

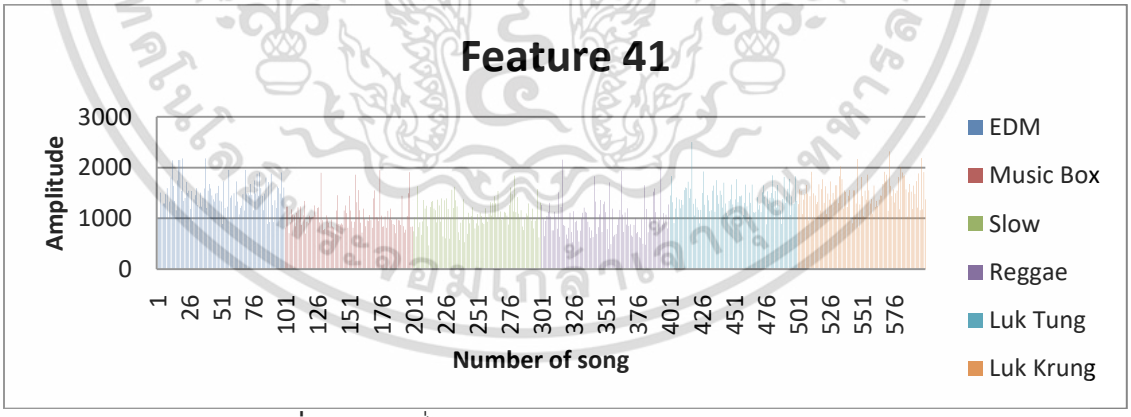
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.22 ค่าที่ได้จากฟีเจอร์ Mean of Pitch

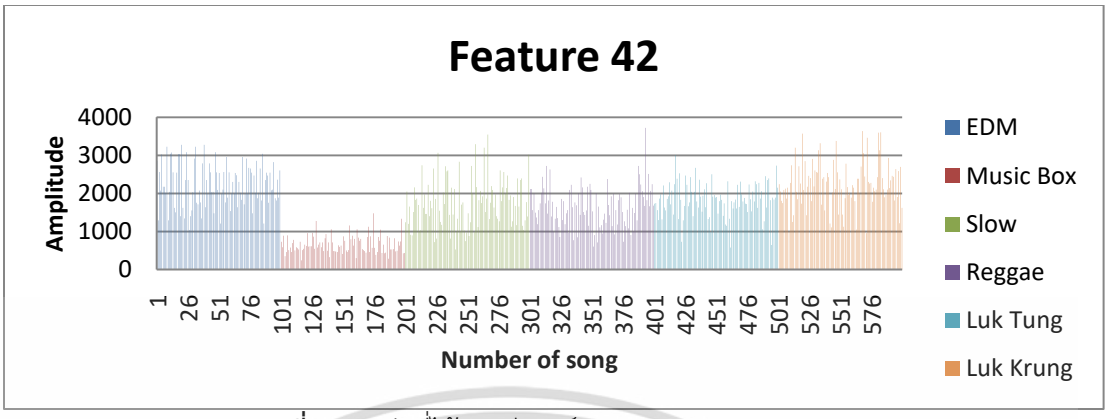


รูปที่ 4.23 ค่าที่ได้จากฟีเจอร์ Std of Pitch

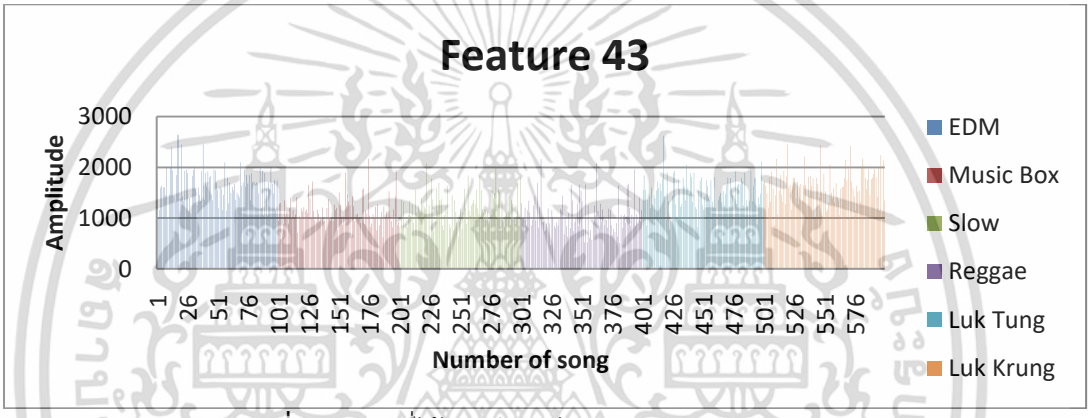


รูปที่ 4.24 ค่าที่ได้จากฟีเจอร์ Mean of Pitch Strength

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.25 ค่าที่ได้จากฟีเจอร์ Std of Pitch Strength



รูปที่ 4.26 ค่าที่ได้จากฟีเจอร์ Sum of Pitch Strength

เนื่องจากเมื่อนำฟีเจอร์ทั้งหมดมาแยกประเภทด้วยตัวจำแนกประเภทต่างกัน 4 ชนิด ตามตารางที่ 4.1 พบว่าค่าความถูกต้องที่ได้มีค่าน้อย อาจเป็นเพราะความซ้ำซ้อนของข้อมูล หรือบางฟีเจอร์มีผลทำให้เกิดความสับสน ทำให้การแยกประเภทมีค่าที่คลาดเคลื่อนไปจากความเป็นจริง เราจึงทำการเลือกฟีเจอร์ที่โดดเด่น และไม่มีความซ้ำซ้อนของข้อมูลมาทำการแยกประเภทใหม่ โดยใช้วิธีการของ PSO เข้ามาช่วยในการคัดเลือกฟีเจอร์ หรือที่เรียกว่า feature selection

ตารางที่ 4.1 แสดงค่าความถูกต้องของการใช้ฟีเจอร์ทั้งหมดนำมาแยกประเภทด้วยตัวแยกประเภทชนิดต่างๆ

Classifier	Accuracy (%)
Random Forest	*80.83*
kNN	*75.00*
SVM	*83.33*
Decision Tree	*69.17*

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

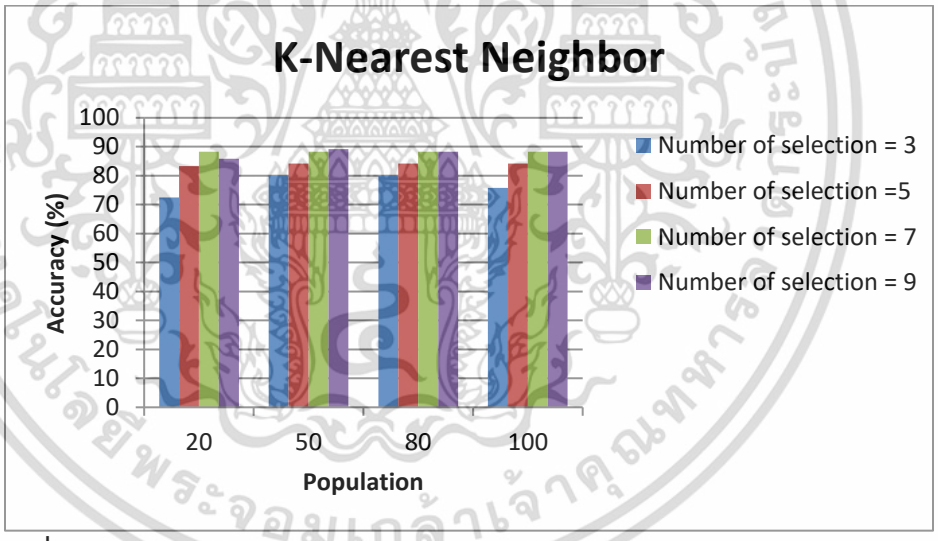
ตารางที่ 4.1 (ต่อ) แสดงค่าความถูกต้องของการใช้ฟีเจอร์ทั้งหมดนำมาแยกประเภทด้วยตัวแยกประเภทชนิดต่างๆ

Classifier	Accuracy (%)
Naïve Bayes	*76.83*

4.2 Feature Selection

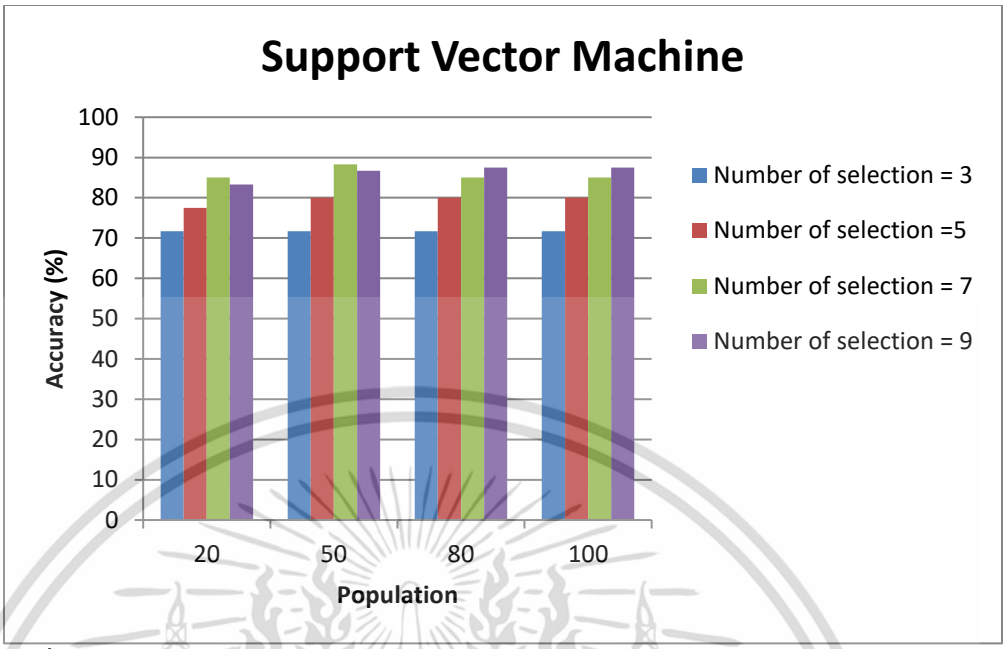
ในขั้นตอนการทำ PSO จะต้องกำหนดสมการ fitness, จำนวน population และจำนวนของฟีเจอร์ที่ต้องการจะเลือกใช้ เราจึงได้ทำการทดลองโดยเลือกใช้สมการ fitness, ค่าของ population และจำนวนของฟีเจอร์ที่ค่าแตกต่างกัน ในส่วนของสมการ fitness ในงานวิจัยนี้ใช้วิธีการของ kNN, SVM และ Decision Tree ส่วนค่า population ที่นำมาใช้ในการทดลองใช้ที่ค่า 20, 50, 80 และ 100 ส่วนจำนวนของฟีเจอร์ที่เลือกใช้ เราได้ทำการทดลองที่ค่า 3, 5, 7 และ 9

รูปข้างล่างแสดงค่าความถูกต้อง โดยใช้ตัวแยกประเภทต่างๆ ในสมการ fitness โดยที่กราฟแนวตั้งคือค่าความถูกต้องที่อ่านได้ กราฟแนวนอนคือจำนวน population ที่ค่าต่างๆ

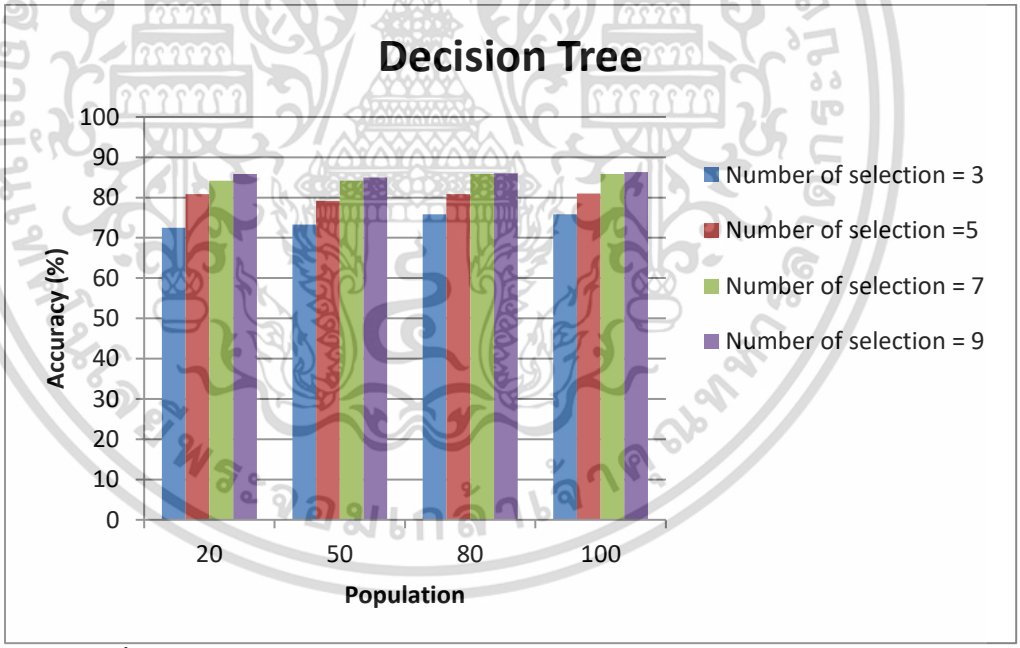


รูปที่ 4.27 ค่าความถูกต้องของสมการ fitness โดยใช้ K-Nearest Neighbor

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.28 ค่าความถูกต้องของสมการ fitness โดยใช้ Support Vector Machine



รูปที่ 4.29 ค่าความถูกต้องของสมการ fitness โดยใช้ Decision Tree

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 แสดงค่าเปรียบเทียบของตัวจำแนกประเภทต่างๆ ค่า populationและจำนวนการเลือกฟีเจอร์ที่ค่าต่างๆ

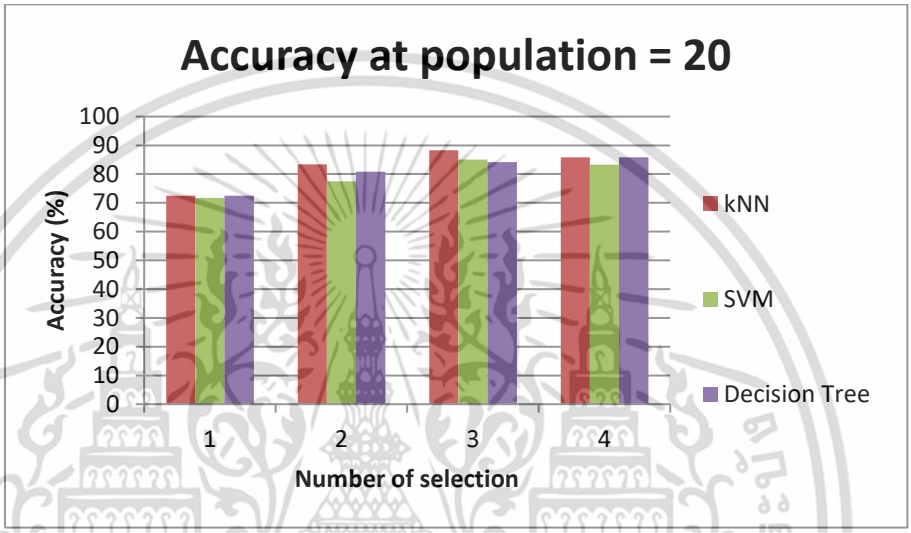
Population	Number of selection	kNN (%)	SVM (%)	Decision Tree (%)
20	3	*72.50*	*71.67*	*72.50*
	5	*83.33*	*77.50*	*80.33*
	7	*88.33*	*85.00*	*84.17*
	9	*85.83*	*83.33*	*85.83*
50	3	*80.00*	*71.67*	*73.33*
	5	*84.17*	*80.00*	*79.17*
	7	*88,33*	*88.33*	*84.17*
	9	*89.17*	*86.67*	*85.00*
80	3	*80.00*	*71.67*	*75.83*
	5	*84.17*	*80.00*	*80.83*
	7	*88.33*	*85.00*	*85.83*
	9	*88.33*	*87.50*	*86.00*
100	3	*75.83*	*71.67*	*75.83*
	5	*84.17*	*80.00*	*81.00*
	7	*88.33*	*85.00*	*85.83*
	9	*88.33*	*87.50*	*86.33*

จากรูปข้างบนเมื่อลองเปรียบเทียบค่าความถูกต้องที่ค่า population ต่างๆ ของแต่ละตัวแยกประเภท พบว่าpopulation ที่ค่า 20 เพียงพอต่อการนำมาใช้งาน เนื่องจากเมื่อ population มีค่าสูงขึ้นค่าความถูกต้องยังคงมีค่าประมาณเดิม ไม่มีการเปลี่ยนแปลงที่มากจนเห็นได้ถึงความแตกต่าง อีกทั้งยังช่วยลดระยะเวลาในการประมวลผลลงไปอีกด้วย เราจึงเลือกใช้ค่า population ที่ 20

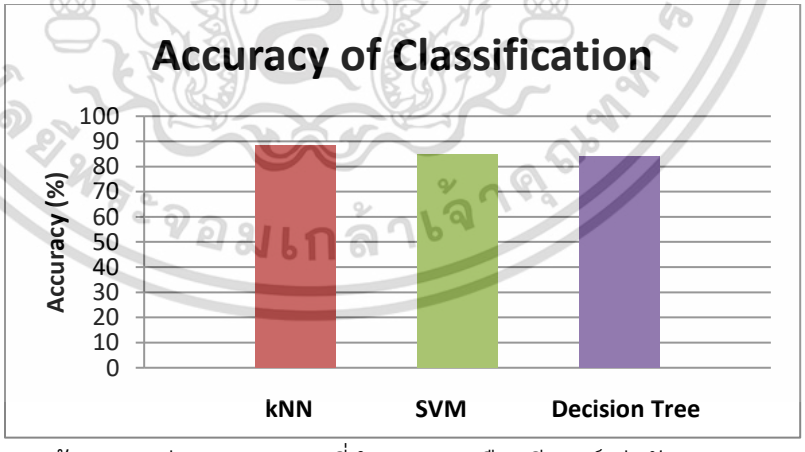
รูปที่ 4.30 เปรียบเทียบค่าความถูกต้องของตัวจำแนกปะเภทต่างๆที่จำนวนการเลือกฟีเจอร์ที่ต่างกัน พบว่าที่จำนวนการเลือกฟีเจอร์จำนวน 7 ฟีเจอร์ ให้ค่าความถูกต้องได้มากที่สุด เมื่อเพิ่มจำนวนการเลือกฟีเจอร์เป็น 9 ค่าความถูกต้องที่ได้มีค่าใกล้เคียงกับกับเลือกจำนวนฟีเจอร์เท่ากับ 7 ดังนั้นเราจึงใช้จำนวนฟีเจอร์ที่เลือก 7 ฟีเจอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 4.31 เปรียบเทียบค่าความถูกต้องของตัวจำแนกปะเภทต่างๆ ที่มีจำนวนการเลือกฟีเจอร์เท่ากับ 7 ฟีเจอร์ และ population เท่ากับ 20 จะเห็นได้ว่าตัวจำแนกประเภท kNN ให้ค่าความถูกต้องมากที่สุด รองลงมาเป็น SVM, Decision Tree และ Random Forest ตามตารางที่ 4.3 ดังนั้นในการกำหนดค่าเริ่มต้นของวิธีการ PSO จึงใช้ตัวจำแนกประเภท kNN เป็นสมการ fitness และมีค่า population เท่ากับ 20 จำนวนฟีเจอร์ที่เลือกใช้เท่ากับ 7 ฟีเจอร์



รูปที่ 4.30 ค่าความถูกต้องของแต่ละ classifier ที่จำนวนการเลือกฟีเจอร์ที่ต่างกัน โดยมีค่า population ที่ 20



รูปที่ 4.31 ค่าความถูกต้องของแต่ละ classifier ที่จำนวนการเลือกฟีเจอร์เท่ากับ 7 และ population เท่ากับ 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

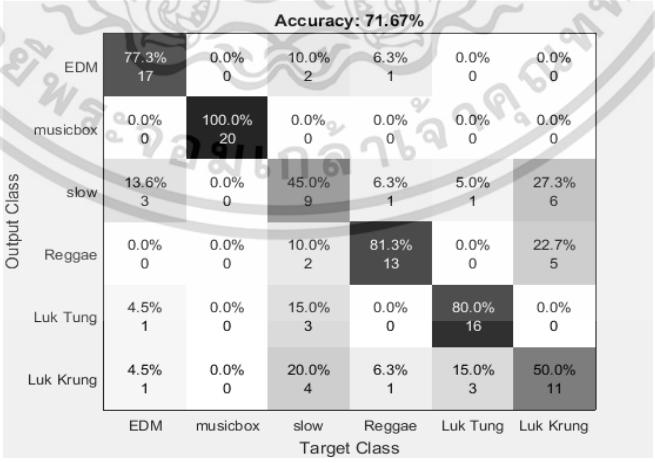
ตารางที่ 4.3 แสดงค่าความถูกต้องของ fitness ที่ตัวจำแนกประเภทต่างๆ ที่ค่า population เท่ากับ 20 และจำนวนการเลือกฟีเจอร์เท่ากับ 7

Classifier	Accuracy (%)
kNN	*88.33*
SVM	*85.00*
Decision Tree	*84.17*

ฟีเจอร์ที่ได้จากการทำ PSO ได้แก่ Strength of Pitch, Std of Pitch, Mean of Spectral Contrast, Std of Spectral Contrast, Std of Volume, Std of Second Mel-Frequency Cepstral Coefficients (MFCC) และ Mean of Zero Crossing

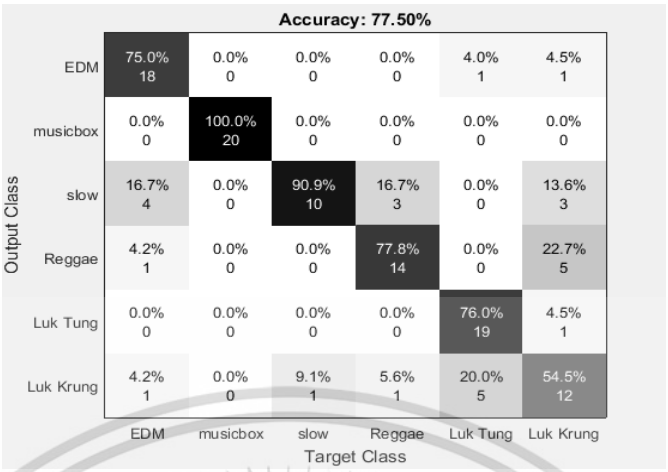
ตารางที่ 4.4 เปรียบเทียบค่าความถูกต้องของตัวจำแนกประเภทต่างๆ ระหว่างใช้ฟีเจอร์ทั้งหมด และใช้ฟีเจอร์ที่เลือกมาจากวิธีการของ PSO

Classifier	All Feature (%)	PSO(%)
Random Forest	*80.83*	*79.17*
kNN	*75.00*	*77.50*
SVM	*83.33*	*84.17*
Decision Tree	*69.17*	*71.67*
Naïve Bayes	*76.83*	*78.33*

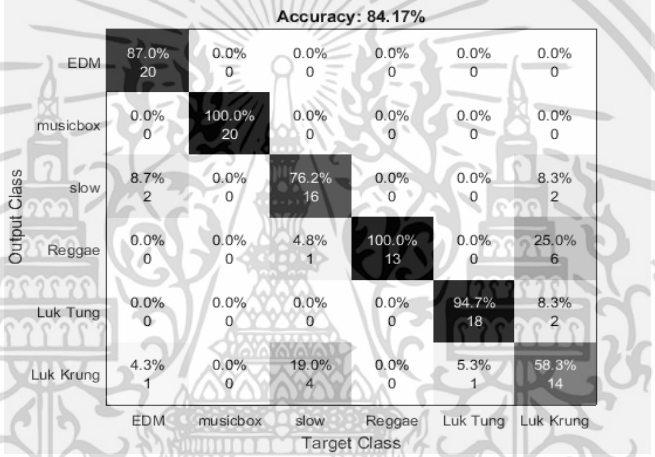


รูปที่ 4.32 Confusion Matrix ของ Decision Tree ที่ผ่านการเลือกใช้ฟีเจอร์ (PSO)

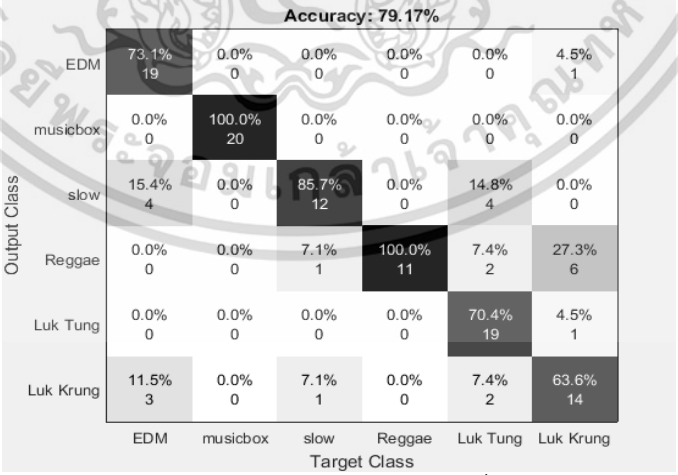
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.33 Confusion Matrix ของ K-Nearest Neighbor ที่ผ่านการเลือกใช้ฟีเจอร์ (PSO)

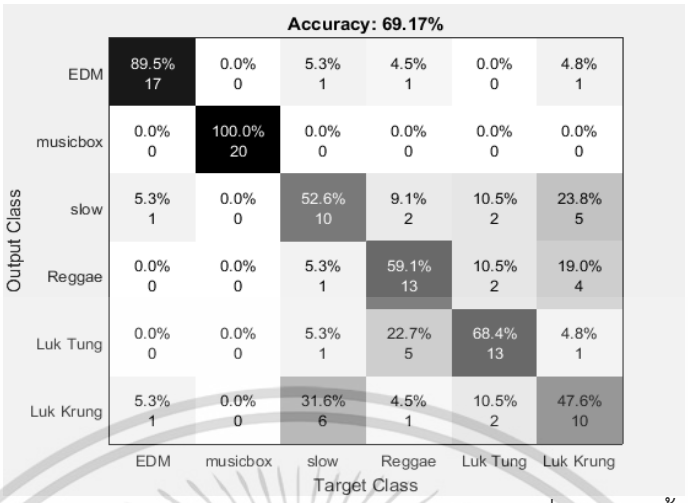


รูปที่ 4.34 Confusion Matrix ของ Support Vector Machine ที่ผ่านการเลือกใช้ฟีเจอร์ (PSO)

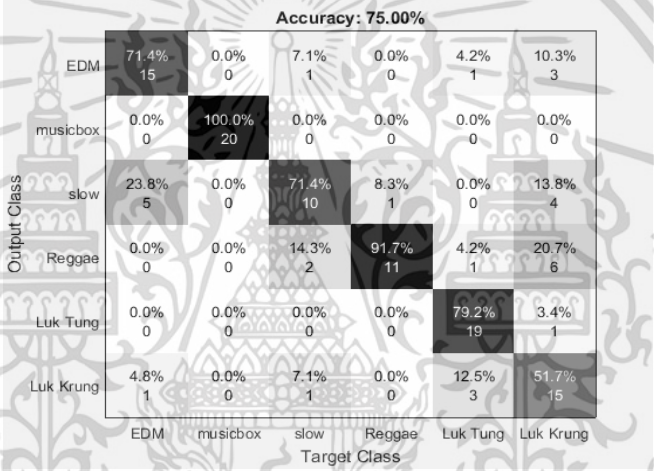


รูปที่ 4.35 Confusion Matrix ของ Random Forest ที่ผ่านการเลือกใช้ฟีเจอร์ (PSO)

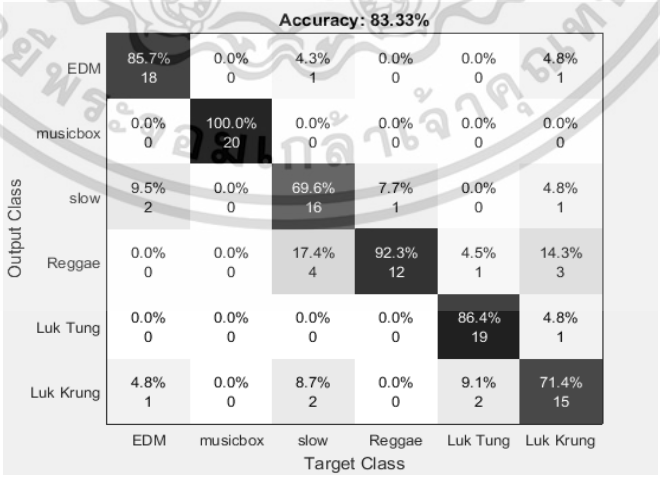
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.36 Confusion Matrix ของ Decision Tree ที่ใช้ฟีเจอร์ทั้งหมด

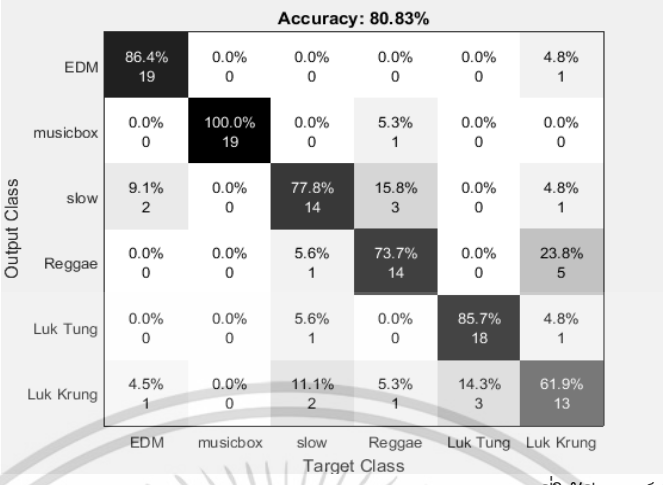


รูปที่ 4.37 Confusion Matrix ของ K-Nearest Neighbor ที่ใช้ฟีเจอร์ทั้งหมด



รูปที่ 4.38 Confusion Matrix ของ Support Vector Machine ที่ใช้ฟีเจอร์ทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.39 Confusion Matrix ของ Random Forest ที่ใช้ฟีเจอร์ทั้งหมด

4.3 Stacking Ensemble

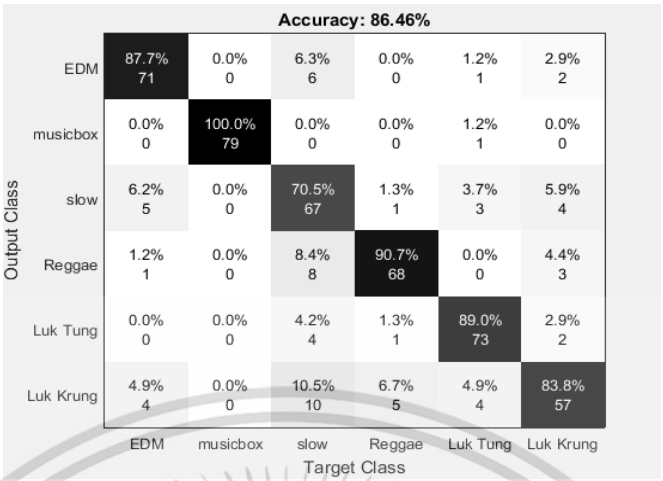
หลังจากขั้นตอนการทำ PSO เพื่อเลือกฟีเจอร์ที่ดีที่สุดมาใช้งาน พบว่าค่าความถูกต้องของแต่ละตัวจำแนกประเภท ยังคงมีประสิทธิภาพต่ำ เราจึงได้ปรับปรุงการทำนาย โดยใช้วิธีการของ Stacking เข้ามาช่วย ซึ่งวิธีการทำ stacking ก็เพื่อรวมโมเดลของตัวจำแนกหลายๆ แบบเข้าไว้ด้วยกัน เพื่อปรับปรุงผลลัพธ์ให้ดียิ่งขึ้น ตัวจำแนกประเภทที่เราจะนำมาสร้างเป็นโมเดลใหม่ ได้แก่ Random Forest, kNN, SVM, Decision Tree และ Naïve Bay

ตารางที่ 4.5 เป็นการทำนายผลลัพธ์ของตัวจำแนกแต่ละประเภทที่ได้จากการทำ 10-fold cross validation เพื่อนำไปใช้เป็นข้อมูลเทรนชุดใหม่ของ Meta Classifier

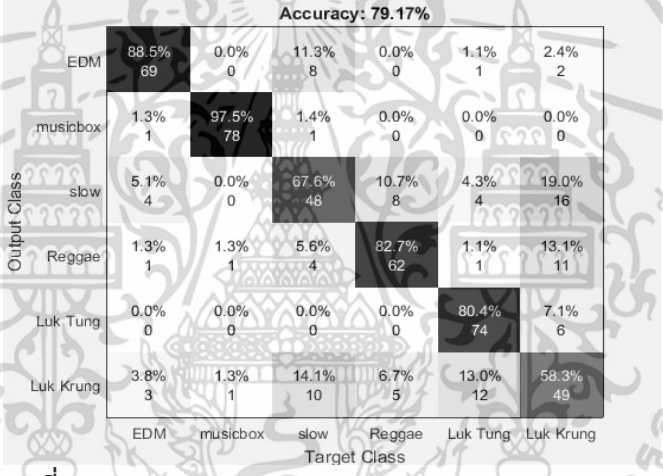
ตารางที่ 4.5 แสดงค่าความถูกต้องของการสร้างโมเดลจากการทำ 10-fold cross validation

Classifier	Accuracy (%)
Decision Tree	*86.46*
kNN	*79.17*
SVM	*83.33*
Random Forest	*88.33*
Naïve Bay	*78.96*

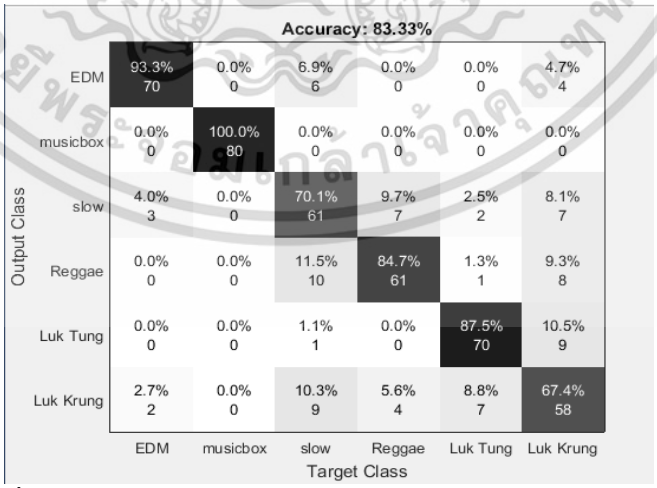
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.40 Confusion Matrix ของ Decision Tree

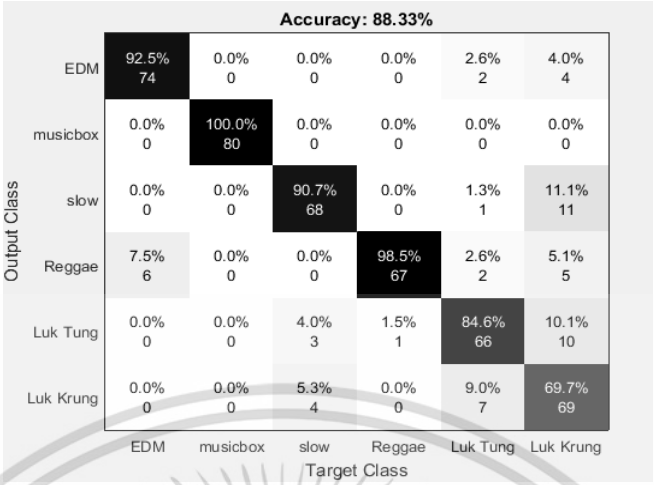


รูปที่ 4.41 Confusion Matrix ของ K-Nearest Neighbor

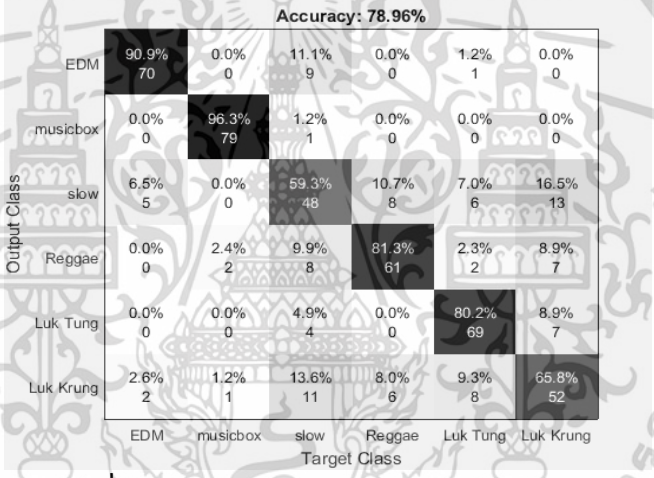


รูปที่ 4.42 Confusion Matrix ของ Support Vector Machine

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



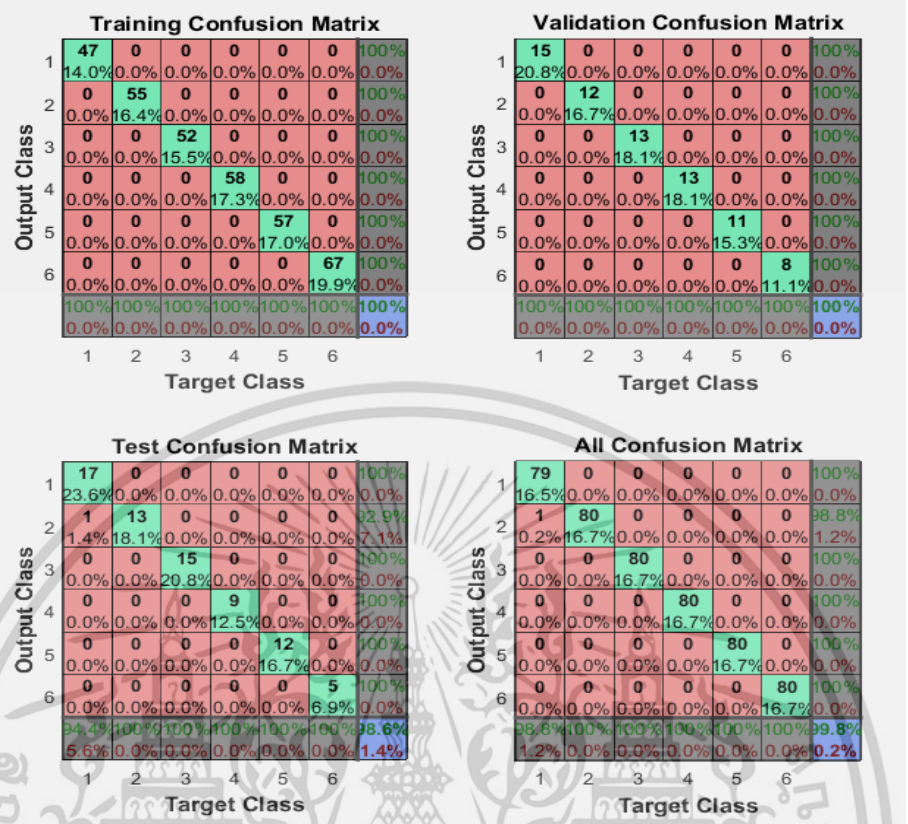
รูปที่ 4.43 Confusion Matrix ของ Random Forest



รูปที่ 4.44 Confusion Matrix ของ Naïve Bay

รูปข้างบน แสดง Confusion Matrix ของตัวจำแนกประเภทต่างๆ ในขั้นตอนของ Base Classifier ตามรูปที่ 3.7 เพื่อนำการทำนายผลมาเข้าเป็นข้อมูลเทรนของชุด Meta Classifier ซึ่งเป็นการรวมตัวจำแนกประเภทหลายๆ ตัวเข้าไว้ด้วยกัน โดยผ่าน Neural Network ในงานวิจัยนี้ใช้รูปแบบของ pattern recognition [6] feed-forward network โดยมีอิทพุทเข้า 5 ใช้ hidden layer 10 ทำการแยกเอาท์พุทออกมา 6 ประเภท ตามรูปที่ 3.11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

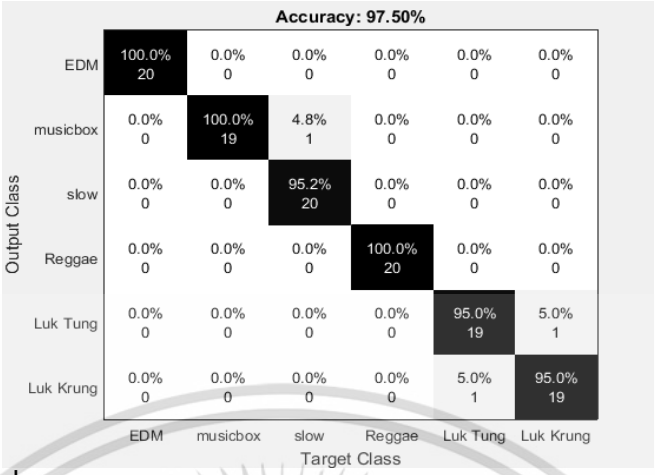


รูปที่ 4.45 Confusion Matrix ของตัวจำแนกประเภท Neural Network แบบ Pattern Recognition

รูปที่ 4.45 คือ Confusion Matrix ของตัวจำแนกประเภท Neural Network แบบ Pattern Recognition โดยช่องสีเขียว คือเปอร์เซ็นต์ของการจำแนกประเภทที่ถูกต้อง และ ช่องสีแดง คือ เปอร์เซ็นต์ของการจำแนกประเภทที่ไม่ถูกต้องและช่องสีฟ้า คือ เปอร์เซ็นต์ของการจำแนกประเภทของโมเดล ซึ่งจากผลการทดลองในรูปที่ 4.45 จะเห็นได้ว่าตัวจำแนกประเภทนั้นมีเปอร์เซ็นต์ความถูกต้องเป็นส่วนใหญ่

เมื่อได้โมเดลใหม่สำหรับการทำนายผลลัพธ์ขั้นสุดท้าย เอาข้อมูลทดสอบมาทดสอบกับโมเดลใหม่นี้ พบว่าเปอร์เซ็นค่าความถูกต้องที่ได้มีค่าเพิ่มมากขึ้นเมื่อเทียบกับการใช้ตัวจำแนกเพียงตัวเดียว ผลลัพธ์ดูได้จากรูปที่ 4.46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.46 Confusion Matrix ของการทำ Stacking Ensemble

เมื่อนำผลลัพธ์จากวิธีของ Stacking มาเปรียบเทียบกับการใช้วิธีอื่นในการจำแนกประเภท โดยใช้ข้อมูลเป็นเพลงไทย จากงานวิจัยต่างๆ พบว่าการทำ Stacking ให้ผลลัพธ์ที่ดีกว่าการจำแนกประเภทอื่นๆ ดูได้ในตารางที่ 4.6

ตารางที่ 4.6 ค่าความถูกต้องของการจำแนกประเภทด้วยวิธีต่างๆ

Title	Feature	Classifier	Accuracy
A Comparison of Audio features of Thai Classical Music Instrument	Chromagram Feature, MFCC Feature, Cepstrum feature	Euclidean Cosine	65.00 % 71.66%
Musical-scale Characteristics for Traditional Thai Music Genre Classification	Chromatic scale, Diatonic Scale, Mode	SVM	84.30 %
Lyric-based Sentiment Polarity Classification of Thai Songs	Lyric Features (unigrams, bigrams and trigrams)	Neural Network Multi-layer perceptron (MLP)	68.00 %

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.6 (ต่อ) ค่าความถูกต้องของการจำแนกประเภทด้วยวิธีต่างๆ

Title	Feature	Classifier	Accuracy
Thai Classical Music Matching Using t-Distribution on Instantaneous Robust Algorithm for Pitch Tracking Framework	Pitch Melody	T - Distribution	99.01 %

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

การวิจารณ์หรืออภิปรายผล (DISCUSSION)

ในงานวิจัยนี้ใช้ชุดข้อมูลเพลงไทยจำนวน 6 ประเภทแนวเพลง ทำการแยกคุณลักษณะของข้อมูลโดยคุณลักษณะที่ใช้ประกอบไปด้วย จังหวะ (Beat related) , เนื้อเสียง (Timbral texture) และ ระดับเสียง (Pitch related) เมื่อทำการแยกคุณลักษณะจะได้ฟีเจอร์ออกมา 43 ฟีเจอร์ จากนั้นนำฟีเจอร์ทั้งหมดไปผ่านขั้นตอนการทำ PSO เพื่อลดความซ้ำซ้อนของข้อมูล โดยเลือกใช้ฟีเจอร์ที่มีคุณลักษณะที่โดนเด่นที่สามารถแยกประเภทของข้อมูลได้อย่างชัดเจนเพื่อทำให้ได้ผลลัพธ์ของการทำนายที่ถูกต้องที่สุด โดยใช้ข้อมูลเทรนจำนวน 360 ข้อมูลทดสอบ 120 หลังจากนั้นทำการทดลองหาค่า population, จำนวนของฟีเจอร์ที่เลือกใช้ และสมการ fitness ที่เหมาะสม เพื่อทำให้ขั้นตอนการเลือกฟีเจอร์มีประสิทธิภาพมากที่สุด ผลลัพธ์จากการทำการทดลองจะได้ค่าที่เหมาะสมของ population คือ 20 จำนวนฟีเจอร์ที่เลือกใช้เท่ากับ 7 ฟีเจอร์ และสมการ fitness ที่เหมาะสม คือสมการของ K-Nearest Neighbor

หลังจากได้ฟีเจอร์ที่ผ่านการทำ PSO นำฟีเจอร์เหล่านี้มาเปรียบเทียบกับฟีเจอร์ทั้งหมดที่ทำการแยกคุณสมบัติทั้งหมด 43 ฟีเจอร์ พบว่าค่าความถูกต้องของฟีเจอร์ที่มาจากการทำ PSO มีค่ามากกว่าการใช้ฟีเจอร์ทั้งหมด โดยที่ตัวจำแนกประเภท Support Vector Machine มีค่าความถูกต้องมากที่สุด 84.17% จากนั้นนำฟีเจอร์ทั้ง 7 ฟีเจอร์มาปรับปรุงปรุงค่าความถูกต้องด้วยวิธีการทำ Stacking Ensemble โดยในชั้นแรกของการทำ Stacking ใช้ตัวจำแนกประเภททั้งหมด 5 ตัวด้วยกัน คือ Random Forest, K-Nearest Neighbor, Support Vector Machine และ Naïve Bay ผลลัพธ์การทำนายที่ได้จากทั้ง 5 ตัวจำแนกประเภทจะถูกนำไปใส่เป็นข้อมูลเทรนในชั้นที่สองของการทำ Stacking ทำการเทรนข้อมูลชุดใหม่ผ่าน Neural Network แบบ Pattern Recognizes จะได้โมเดลใหม่สำหรับการทำนายข้อมูลทดสอบ เมื่อนำโมเดลใหม่มาทดสอบกับข้อมูลทดสอบจะได้ค่าความถูกต้อง 97.50% ซึ่งถือว่ามีความถูกต้องมากกว่าการใช้ตัวจำแนกประเภทแค่เพียงตัวเดียว ดังนั้น การทำ Stacking ถือได้ว่ามีประสิทธิภาพมากในการปรับปรุงค่าความถูกต้องของการทำนายข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

[1] G. Tzanetakis and P. Cook, Musical genre classification of audio signals, IEEE Transactions on Speech and Audio Processing 10, 2002, pp.293–302.

[2] Max Chevalier and Christine Julien, Music recommendation, in collaborative and social information retrieval and access, 2009, pp.212.

[3] T. Tolonen and M. Karjalainen, A computationally efficient multipitch analysis model, IEEE Transactions Speech Audio Processing, vol. 8, Nov. 2000, pp.708–716.

[4] M. Tamura, T. Masuko, K. Tokuda, and T. Kobayashi, Adaptation of pitch and spectrum for HMM-based speech synthesis using MLLR, IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 1, May 2010, pp. 805–808.

[5] M. Dash and H. Liu, Feature selection for classification, Intelligent Data Analysis, vol. 1, no. 4, 1997, pp. 131–156.

[6] Jayanta Kumar Basu, Debnath Bhattacharyya, Taihoon Kim, Use of Artificial Neural Network in Pattern Recognition, Inter-national Journal of Software Engineering and Its Applications vol. 4, no.2, April 2010, pp.24-34.

[7] Piero Baraldi, Enrico Zio, Davide Roverso, Feature Selection For Nuclear Transient Diagnostics, OECD Halden Reactor Pro-ject, January 2007.

[8] Beatriz A. Garro and Roberto A. Vazquez. Designing Artificial Neural Networks Using Particle Swarm Optimization Algorithms, Computational Intelligence and Neuroscience, 2015.

[9] Bing Xue, Classification Algorithms in Particle Swarm Optimization for Feature Selection in Classification, 2014, pp.21-25.

[10] David Opitz and Richard Maclin, Popular Ensemble Methods: An Empirical Study, Journal of Artificial Intelligence Research 11, 1999, pp. 169-198.

[11] Kristína Machová, František Barčák and Peter Bednár, A Bagging Method using Decision Trees in the Role of Base Classifi-ers, Acta Polytechnica Hungarica, vol. 3, no. 2, 2006.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

[12] Maneesh Singhal and Ramashankar Sharma, Optimization of Naïve Bayes Data Mining Classification Algorithm, International journal for research in applied science and engineering technology (IJRASET), vol. 2 Issue VIII, August 2014.

[13] L. Oliveira, R. Sabourin, F. Bortolozzi, and C. Suen, Feature selection using multi-objective genetic algorithms for handwritten digit recognition, in 16th International Conference on Pat-tern Recognition (ICPR 02), vol. 1, 2002, pp.568–571.

[14] Jun Yang, Fa-Long Luo, Arye Nehorai, Spectral contrast enhancement: Algorithms and comparisons, Speech Communication 39, 2003, pp.33–46.

[15] Jayanta Kumar Basu, Debnath Bhattacharyya, Taihoon Kim, Use of Artificial Neural Network in Pattern Recognition, Inter-national Journal of Software Engineering and Its Applications vol. 4, no.2, April 2010, pp.24-34.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สูตรต่างๆ ที่ใช้ในการหาฟีเจอร์

Feature	Equation
Spectral Centroid	$C = \frac{\sum_{n=1}^{N} M_t[n] * n}{\sum_{n=1}^{N} M_t[n]}$
Spectral Flux	$F = \sum_{n=1}^{N} (N_t[n] - N_{t-1}[n])^2$
Spectral Rolloff	$\sum_{n=1}^{R_t} M_t[n] = r * \sum_{n=1}^{N} M_t[n]$
Volume	$v(n) = \sqrt{\frac{1}{N} \sum_{t=0}^{N-1} S_n^2(i)}$
Zero Crossing	$\sum_{n=1}^{N} \lvert sign(x[n]) - sign(x[n-1]) \rvert$
Short Time Energy	$E_{\hat{n}} = \sum_{m=-\infty}^{\infty} (x[m] w[\hat{n} - m])^2$
Energy	$\sum_{n} \lvert x(n) \rvert^2$
Low Energy	$\sqrt{\frac{1}{N} \sum_{n} \lvert x(n) \rvert^2}$
Mel-Frequency Cepstral Coefficients	$m = 2595 log_{10}(\frac{f}{700} + 1)$
Period Peak Unfolded of Pitch (Pitch)	$n = 12 log_2 \frac{f}{440} + 69$
Period Peak Folded of Pitch	$c = n \bmod 12$
Beat Per Minute (BPM)	$BPM = \frac{count}{t} * 60$
Ratio Amplitude Histogram of Beat	$\frac{Second\ Histogram}{First\ Histogram}$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สูตรต่างๆ ที่ใช้ในการหาฟีเจอร์ (ต่อ)

Spectral Contrast	$$\boldsymbol{Peak_k = log\left\{\frac{1}{aN}\sum_{i=1}^{aN}\acute{X}_{k,i}\right\}}$$ $$\boldsymbol{Valley_k = log\left\{\frac{1}{aN}\sum_{i=1}^{aN}\acute{X}_{k,N-i+1}\right\}}$$ $$\boldsymbol{SC_k = Peak_k - Valley_k}$$
Spectral Mean	$$\overline{X_k} = \frac{1}{N}\sum_{i=1}^{N}X_{k,1}$$

โค้ด MATLAB โดยใช้วิธีการ PSO

```
% Boundary
cloumn=size(Data,2);
if cloumn<=nOfSelection
   error('Number of Selection must be positive and less than Selection boundary')
end
Boundary=[1,cloumn];

% Initialize Constants
weight=0.2;
constant1=2;
constant2=2;

% Initialize Pariticles
Positions=zeros(population,nOfSelection);
for i=1:population
    tempVar = randperm(Boundary(2));
    Positions(i,:) = tempVar(1:nOfSelection);
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
% velocity initialization
Velocity = ones(population,nOfSelection);

% Initialize Global best
itration = 1;
globalBest.value(itration) = inf;
%% Optimizing
Particles = struct;

while(1)
% Calculating Fitness Values
    for i=1:population
        Particles(itration).position(i,:) = Positions(i,:);
        X = Data(:,Positions(i,:));
        ctree = fitctree(X,YY); % create classification tree
        an = predict(ctree,testdata(:,Positions(i,:)));
        C = confusionmat(targets,an);
        CC = C(1,1)+C(2,2)+C(3,3)+C(4,4)+C(5,5)+C(6,6);
        Particles(itration).value(i) =(CC/120)*100;
        Acc(i,itration) = (CC/120)*100;
    end
    [pointvalue,point] = max(Acc(:,itration));
    Point(itration,:)= point;
    PositionAcc(itration,:) = Positions(Point(itration,:),:);

    % Update Position Best
    [Particlesbest.value(itration),ind] = max(Particles(itration).value);
    Particlesbest.position(itration,:) = Particles(itration).position(ind,:);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
% Update Global Best
    if globalBest.value(itration) > Particlesbest.value(itration)
 [globalBest.value(itration),globalBest.value(itration+1)]=deal(Particlesbest.value(itration));
        globalBest.position(itration,:) = Positions(ind(1),:);
    else
        globalBest.value(itration+1) = globalBest.value(itration);
        globalBest.position(itration,:) = globalBest.position(itration-1,:);
    end
    fprintf('\nIteration %d gbest value of  fitness= %f',itration,globalBest.value(itration))

 % Velocity Update
Velocity=(weight*Velocity)+(constant1*rand(1)*(repmat(Particlesbest.position(itration,:),po
pulation,1)-Positions))+(constant2*rand(1)*(repmat(globalBest.position(itration,:),
population,1)-Positions));

% Postion Update
Positions = Positions+round(Velocity);

% Boundary Checking for Position
    Positions(Positions>Boundary(2)) = round(rand(1)*(Boundary(2)-1))+1;
    Positions(Positions<Boundary(1)) = round(rand(1)*(Boundary(2)-1))+1;
    for i=1:population
        if length(unique( Positions(i,:) )) ~= nOfSelection
        tempVar = randperm(Boundary(2));
        Positions(i,:) = tempVar(1:nOfSelection);
        end
    end

% Loop Breaking is Positions all are same or itertion acheived maximum
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
% itertion
    count=0;
    for i=1:population-1;
        if Positions(i,:) == Positions(i+1,:)
            count = count+1;
        end
    end
    if (count == population-1) || (itration >= itrationMax)
        fprintf('\n******Iteration completed******\n')
        break
    end
     itration = itration+1;
end
Selection=globalBest.position(itration,:);
SelectionValue=globalBest.value(itration);
```

โค้ด MATLAB โดยใช้วิธีการ DT

```
%Decistion tree
ctree = fitctree(a,traindatalabel); % create classification tree
DT = predict(ctree,te);
```

โค้ด MATLAB โดยใช้วิธีการ kNN

```
%k nearest neighbors
Mdl = fitcknn(a,traindatalabel,'NumNeighbors',20,'Standardize',1);
kNN = predict(Mdl,te);
```

โค้ด MATLAB โดยใช้วิธีการ SVM

```
%SVM
traindatalabelsvm = categorical(traindatalabel);
classOrder = unique(traindatalabelsvm);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
rng(1); % For reproducibility
t = templateSVM('Standardize',1);
CVMdl = fitcecoc(a,traindatalabelsvm,'Holdout',0.30,'Learners',t,'ClassNames',classOrder);
CMdl = CVMdl.Trained{1}; % Extract trained, compact classifier
SVM = predict(CMdl,te);
```

โค้ด MATLAB โดยใช้วิธีการ Random Forest

```
%Random forest
BaggedEnsemble = generic_random_forests(a,traindatalabel,2000,'classification');
RF = predict(BaggedEnsemble,te);
```

โค้ด MATLAB โดยใช้วิธีการ Naïve Bayes

```
%Naive Bayes
NModel1 = NaiveBayes.fit(traindata,traindatalabel);
NB = predict(NModel1,traindata);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Music Genre Classification of audio signals Using Particle Swarm Optimization and Stacking Ensemble

Krittika Leartpantulak and Yuttana Kitjaidure
Department of Electronic Engineering, Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang
Bangkok, 10520, Thailand
E-mail: Leartkrittika@outlook.com

***Abstract*—Genre classification is a process of grouping similarities, such as patterns, styles, or objectives with management data as already in the music (e.g. pop and rock). It is used along with the classification of topics.**

This paper will classify songs from audio signal to a hierarchy of musical genre by using feature extraction. Trimbral texture, rhythmic content and pitch content are used as the main feature sets. Feature selection is selected by using Particle Swarm Optimization (PSO) and sent selected feature to classification. The result in classification has low accuracy. Thus, using stacking ensemble method is to improve the prediction.

In this paper, the purpose is to improve the prediction by using stacking ensemble method. Stacking ensemble that have the second level is base classifier and meta-classifier. In base classifier consists of 5 classification; K-Nearest Neighbors (k-NN), Decision Tree (DT), Random Forest, Support Vector Machines (SVM); and Naïve Bayes. This process is generated to build multiple classifier predictors and sent it to meta-classifier. In the process of meta-classifier will create new model to predict test data. The new model has been created from neural network which train data is the output of base classifier.

***Index Terms*—Genre, Classification, PSO, Feature selection, Stacking.**

I. Introduction

The music genres are category descriptions that use to customize the music. The classification of music into the genre, the perception of texture and rhythmic could be described characteristic the particular structure of the music. Human has distinguished [7] in a classification of music as investigated in [6] which shows that human can accurately predict musical genres based on 250 milliseconds of audio. Therefore, techniques for automatic genre classification are useful for developing information music retrieval systems.

George Tzanetakis and Perry Cook were among the first researchers who write on the topic of musical genre classification. The features set for timbral texture, rhythmic content and pitch content are proposed [1].

In PSO, each particle solution of the problem is represented as a particle, which is encoded by a vector or an array. The particles move in the searching space to find the best of optimal solutions. During each movement, a particle can remember the best experience. The particle swarm searches for the best solution by updating the position of each particle based on the best experience of its own and its neighboring particles [12].

This paper is structured as follows. A dataset uses in section 2. Feature extraction and specific feature set for rhythmic structure, timbral texture and pitch content of musical signals are described in section 3. Section 4 will describe feature selection by using PSO algorithm, then categorize a music genre for classification and stacking ensemble to improve prediction, described in section 5 and section 6 is a conclusions and future directions.

II. Data Set

In this paper, we use a Thai music collection as the audio. Each data set consists of 600 audio tracks 60 seconds length. It contains 6 genres; EDM, music box, slow, reggae, city, and country. Each genre is represented by 100 tracks.!!

III. Feature Extraction

Feature extraction involves a reducing amount of data required to describe a large data sets by using group classification such as pitch, the intensity of music or beat is for extracting characteristic of each music.

In this paper, we use the main three features sets for feature extraction implements the original feature set proposed by Tzanetakis & Cook [1]. The features can be split into three groups; beat related, timbral texture and pitch related.

A. Rhythmic Content Features (Beat)

The rhythmic content feature indicates the movement of the music signal over a certain period of time including information such as the consistency of rhythm.! Rhythmic content features calculated by beat histogram calculation are resulting in 5 dimensional feature vector are beat per minute (bmp), first histogram peak, second histogram peak, ratio amplitude histogram, and beat strength [1].

B. Timbral Texture Features

Timbral features are often used for music speech and speech recognition by distinction a variety of sounds with the same rhythmic or similar content [2]. Timbral features consist of low energy, mean and standard deviation (std) of spectral centroid, spectral flux, spectral roll off, spectral contrast, spectral mean, frequency histogram [14], volume, zero-crossing, energy, first mel-frequency cepstral coefficients (mfcc), second mfcc. Third mfcc, fourth mfcc, fifth mfcc [1].

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

C. Pitch Content Features

The pitch content features [4], the multiple pitch detection algorithms described by Tolonen and Karjalainen [3] are utilized. These features consider the subjective assessment of pitch frequency, statistical analysis of pitch mean and standard deviation. The features of pitch are all strength of pitch, sum of pitch strength, period peak unfolded, period peak unfolded, amplitude maximum folded [1], mean, standard of pitch, and pitch strength.

The final feature vector is 43-dimensional (beat: 5; timbral texture: 29; pitch: 9).

IV. FEATURE SELECTION

In classification, the features can describe the particular characteristic in the datasets. Feature selection is to select distinguished feature from overall feature for making an obvious classification. Unrelated or redundant features may reduce the classification efficiency [5]. Thus, feature selection is proposed as a data pre-processing step to reduce irrelevant features. There are several methods to find feature selection. In this paper, we use PSO because it is easy to implement and work efficiently for solving problems continuously, but the disadvantage is. It will converge to premature convergence.

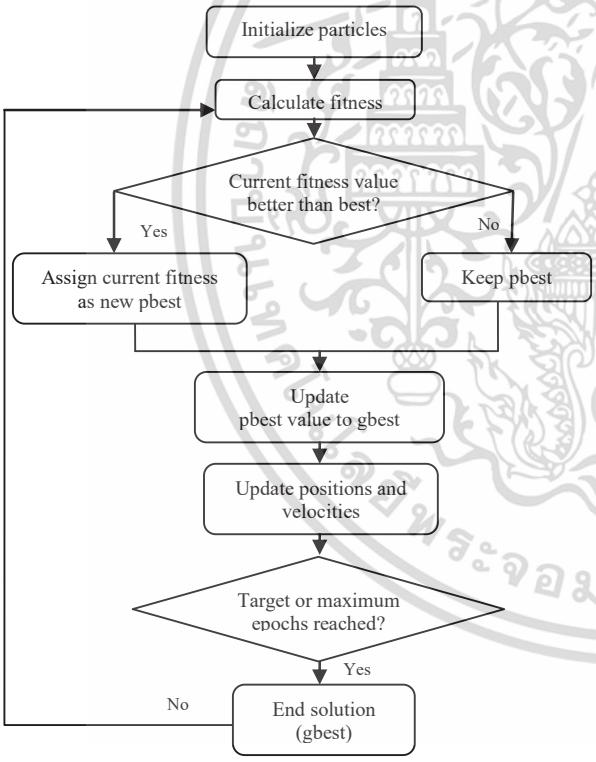


Fig. 1. Flow chart of PSO

A. Particle Swarm Optimization (PSO)

Particle swarm optimization [8] is a swarm intelligence algorithm inspired by the social behavior of birds flocking or fish schooling. PSO algorithm is inspired on the movements of the best member of the population by using the set of solutions moving in a search space with the aim to achieve the best position or solution.

The population is considered as a group of particles i where each represents a position x_i, $i = 1, 2,\ldots, M$ in a multidimensional space. These particles are evaluated in a particular optimization function to recognize their fitness value and collect the best solution. All particles change the position in the search area according to a velocity function $\mathbf{v}_i$ which consider the best position of a particle in a population p_g as well as their own best position p_i. The particles will move in each iteration to different position until they reach an optimum position. At each time t, the particle velocity i is updated by using equation 1.

$$V_i(\,t+1) = wv_i(\,t\,)+c_1r_1(p_i(\,t\,)-x_i(\,t\,))+c_2r_2(\,p_g(\,t\,)-x_i(\,t\,)) \quad (1)$$

where w is the inertia weight and normally set up from 1 to 0 during an iteration run, c_1 and c_2 are acceleration coefficients, generally the value is 2 but can use other values between 0 to 4, r_1 and r_2 are uniformly distributed random numbers between 0 and 1. Updating velocity enables the particle i to search for its best individual position $p_i(t)$, and the best global particle position i is computed as

$$x_i(\,t+1\,) = x_i(\,t\,) + v_i(\,t+1\,) \quad (2)$$

In the process of PSO we have many model selections (random forest, kNN, SVM and DT) to find fitness value called wrapper [16]. Wrapper methods is a feature selection with modeling based on the set of features that are defined and measuring the performance of the model then choose the set of features that make model to the most effective, so we use method for finding the best model to fitness function.

In this process separate dataset to train data 360 and test data 120 to find the best fitness value by using fitness equation with random forest, kNN, SVM and DT. In PSO is use continuous optimization problems and set initialize value of c_1, c_2 are 2, w is 0.2

V. CLASSIFICATION

Classification manages objects in a group with the same category and the same characterize, which have many different classification algorithms in machine learning. Six popular were use in this process are K-nearest neighbor (KNN), Decision Tree classifiers (DT), Random Forest, Support Vector Machine (SVM), Naive Bayes and Neural Classifiers.

A. K-Nearest Neighbors (k-NN)

The k-neighbor neighbor classifier is the type of learning algorithm [9] by using the calculated distance between samples in the test suite and every case in the training set.

B. Decision tree (DT)

DT learning is an algorithm to estimate discrete-valued functions [9], created by selecting the best feature to extract training data and expand the leaf node of tree until reaching the threshold.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

C. Random Forest

The Random forest algorithm is a machine learning technique by randomly selected attributes from many different sets and creates a model with several decision tree techniques.

D. Support Vector Machines (SVM)

SVM is a popular machine learning method [9]. This method is based on the concept of decision planes that define decision boundaries. This paper applies multi-class SVM by using the Error Correcting Output Codes (ECOC) [13] method.

E. Naïve Bayes

Naïve Bayes Classifier uses Bayesian theory [12]. Naive Bayes classifier represents each class of probability and finds the most class likely classify.

F. Neural Network (ANN)

An Artificial Neural Network (ANN) is an information processing pattern that is inspired by the way biological nervous systems, such as the brain, process information.

The ensemble learning methods are used for classification improvement. For summing model of each classified to get the best accuracy called decision fusion in this use stacking ensemble method.

Ensemble methods are meta-algorithms that combine several machine learning techniques into one predictive model, three methods for creating ensembles are bagging [11], boosting [10] and stacking, in this paper, we use stacking to create an ensemble.

Stacking is a class of algorithms that involves training a second-level, it called meta-learner, to find the optimal combination of the base learners, unlike bagging and boosting. In process of stacking ensemble use kNN, DT, Random Forest, SVM and Naïve Bayes for the base classifier and use the neural network for meta-classifier, it is show in fig. 2.

In process of base classifier (level-0 model) are built the multiple classifier systems by using k-fold cross validation on a single dataset. The predictions of level-0 classifiers represent the attributes in a new training set and fed into a meta-classifier (level-1 data), which keep the original class labels. Stacking utilizes a meta-learner (level-1 model) is to create a model for predictions final decision.

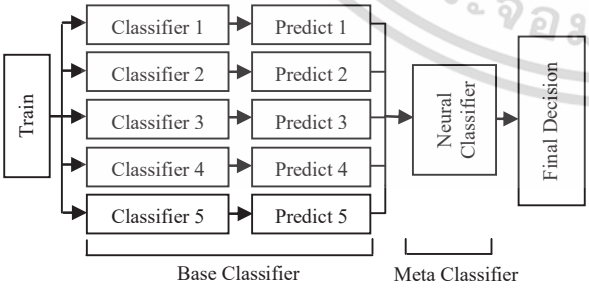


Fig. 2. Block Diagram of Stacking Ensemble

VI. RESULT AND CONCLUSION

In this paper, we use Thai music for dataset, which has 6 genres are EDM, music box, slow, reggae, city, country and separate feature from dataset by using beat, timbral texture and pitch related, resulting in 43-dimensional feature vector. Then use PSO to select features with fitness function which is kNN classifier. The result from feature selected is 7 features and then sent it to a stacking ensemble following the fig. 3.

In PSO, use population and number of selections to find appropriate value to give maximum accuracy by using population which is 20, 50, 80 and 100. Then the numbers of selection used values which are 3, 5, 7, and 9. When population values and numbers of selection is increased, the accuracy has no difference from previous value. Thus, appropriate value of population is 20. Number of selection value is 7, it is suitable to make maximum accuracy, it has shown in fig. 4.

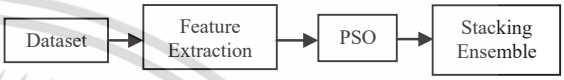


Fig. 3. Block Diagram of music genre classification

The best model of classification (wrapper method) to find fitness function is kNN, which accuracy value is 88.33%, the accuracy value of SVM is 85%, the accuracy value of DT is 84.17% and the accuracy value of Random forest is 80.83%, thus use kNN to fitness function of PSO. Table I shows accuracy values of classified in PSO by using population by 20 and 7 is the number of selections to find the best fitness value.

TABLE I. ACCURACY OF MODEL SELECTION IN PSO FITNESS VALUE

Classified	*Accuracy (%)*
Random Forest	83.33%
kNN	88.33%
SVM	85.00%
DT	84.17%

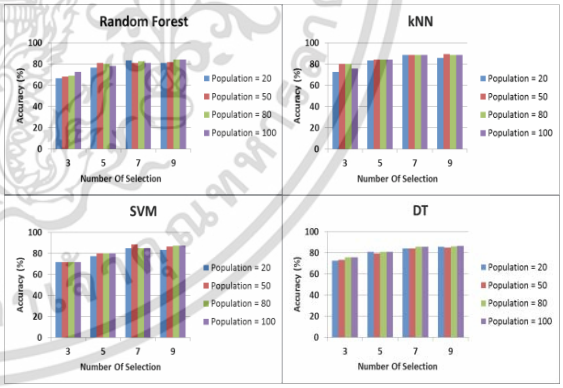


Fig. 4. Population and number of selections of classification

The result of PSO is feature selected and sent it to classification, which has trained data 480 and test data 120. In classification process found SVM has maximum accuracy follow table II. The feature selected used in this classified are strength of pitch, standard deviation of pitch, mean of spectral contrast, standard deviation of spectral contrast, standard deviation of volume, standard deviation of second mel-frequency cepstral coefficients (MFCC) and mean of zero-crossing.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

TABLE II. ACCURACY OF CLASSIFICATION BETWEEN ALL FEATURE AND FEATURE SELECTION

Classified	*Accuracy of all feature*	*Accuracy of features selection*
Random Forest	80.83%	79.17%
kNN	75.00%	77.50%
SVM	83.33%	84.17%
DT	69.17%	71.67%

Table II shows the accuracy values of classified by using feature selected as mentioned above and performance of all feature, it can see the accuracy of PSO feature selection better than use all features. The classifications that use in feature selected are random forest, kNN, SVM and DT to classify. SVM is the best classifier, its accuracy is 84.17%.

In a confusion matrix, the columns correspond to the actual genre and the row to predict genre. The accuracy from Table II is low. Thus, we use stacking ensemble method to improve accuracy. In stacking ensemble, training model to train the data of each classifier to select base classifier by using k-fold cross validation and prediction from base classifier fed into meta-classifier for creation the new model to predict the final decision. Using neural network in meta-classifier has accuracy 97.50% following the figure. 5. shown the confusion accuracy matrix of stacking ensemble. The model of neural network is pattern recognition [15] with feed-forward network. Train dataset of neural model is prediction from each classifier in base classifier following fig.2. The final decision model is created from training model of neural and used the final model to predict test data.

In base classifier process, use 10-fold cross validation by dividing data train 9, data test 1, after that, make training model of each classifier for finding the best accuracy by kNN classifier using k equals 20. Random Forest classifier use trees equals 2000, SVM used to multi-classifier by using ECOC. Then compare of each fold of each classifier to find the best accuracy, use that fold to model for predicting training. Test dataset and sent them to meta-classifier. In meta-classifier use neural data train 70%, data test 15%, validation 15%, which has the best input of each classifier from based classifier, then sent to hidden layer with the number of hidden layers equals 10. Finally, we receive the new model to predict test data.

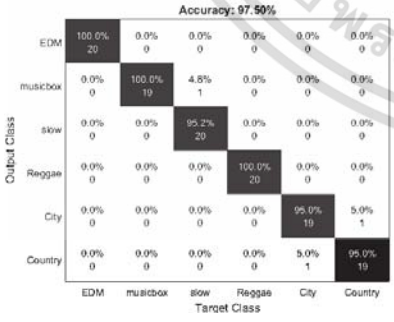


Fig. 5. Confusion accuracy matrix of stacking ensemble

The difference between table I and table II are that table I compared accuracy of each classifier in PSO for find the best fitness classifier and use them as fitness value in PSO. Table II is comparison accuracy between using all feature with feature selected.

When analyzing the data, it was found that the use of stacking ensemble and feature selection help to increase the accuracy of the prediction. Compared to using all of feature, because all features might have duplication, it makes the data confused, so it has less accurate than select some feature.

REFERENCES

[1] G. Tzanetakis and P. Cook, Musical genre classification of audio signals, IEEE Transactions on Speech and Audio Processing 10, 2002, pp.293–302.

[2] Max Chevalier and Christine Julien, Music recommendation, in collaborative and social information retrieval and access, 2009, pp.212.

[3] T. Tolonen and M. Karjalainen, A computationally efficient multipitch analysis model, IEEE Transactions Speech Audio Processing, vol. 8, Nov. 2000, pp.708–716.

[4] M. Tamura, T. Masuko, K. Tokuda, and T. Kobayashi, Adaptation of pitch and spectrum for HMM-based speech synthesis using MLLR, IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 1, May 2010, pp. 805–808.

[5] M. Dash and H. Liu, Feature selection for classification, Intelligent Data Analysis, vol. 1, no. 4, 1997, pp. 131–156.

[6] D. Perrort and G. Jerdigen, An exploration of factors in the identification of musical style, in proceeding of the 1999 society for music perception and cognition, pp.88.

[7] Martin, K.,D., Scheirer, E.D. and Vercoe, B., L., "Musical content analysis through models of audition,"in Proceedings of the 1998 ACM Multimedia Workshop on Content-Based Processing of Music.

[8] Beatriz A. Garro and Roberto A. Vazquez. Designing Artificial Neural Networks Using Particle Swarm Optimization Algorithms, Computational Intelligence and Neuroscience, 2015.

[9] Bing Xue, Classification Algorithms in Particle Swarm Optimisation for Feature Selection in Classification, 2014, pp.21-25.

[10] David Opitz and Richard Maclin, Popular Ensemble Methods: An Empirical Study, Journal of Artificial Intelligence Research 11, 1999, pp. 169-198.

[11] Kristína Machová, František Barčák and Peter Bednár, A Bagging Method using Decision Trees in the Role of Base Classifiers, Acta Polytechnica Hungarica, ! vol. 3, no. 2, 2006.

[12] Maneesh Singhal and Ramashankar Sharma, Optimization of Naïve Bayes Data Mining Classification Algorithm, International journal for reseach in applied science and engineering technology (IJRASET), vol. 2 Issue VIII, August 2014.

[13] L. Oliveira, R. Sabourin, F. Bortolozzi, and C. Suen, Feature selection using multi-objective genetic algorithms for handwritten digit recognition, in 16th International Conference on Pattern Recognition (ICPR 02), vol. 1, 2002, pp.568–571.

[14] Jun Yang , Fa-Long Luo ,Arye Nehorai, Spectral contrast enhancement: Algorithms and comparisons, Speech Communication 39, 2003, pp.33–46.

[15] Jayanta Kumar Basu, Debnath Bhattacharyya, Tai-hoon Kim, Use of Artificial Neural Network in Pattern Recognition, International Journal of Software Engineering and Its Applications vol. 4, no.2, April 2010, pp.24-34.

[16] Piero Baraldi, Enrico Zio , Davide Roverso , Feature Selection For Nuclear Transient Diagnostics, OECD Halden Reactor Project, January 2007.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ-นามสกุล	นางสาวกฤติกา เลิศพันธุลักษณ์
วัน เดือน ปีเกิด	14 พฤศจิกายน 2537 ที่กรุงเทพมหานคร
ประวัติการศึกษา	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ความชำนาญเฉพาะด้าน	1.) การวิเคราะห์ข้อมูลเสียง
	2.) Algorithm Machine Learning

ประสบการณ์การทางานและผลงานวิจัย

พ.ศ.2562	ผลงานวิจัย Music Genre Classification based on audio signals Using Particle Swarm Optimization and Stacking Ensemble

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้