

An analysis of Thai music from Spotify mood-based playlists using machine learning methods

Final presentation

27 April 2022

รายชื่อนิสิต

นาย ณัฐดนัย วีระถาวร

61102010143

นาย ภฤศ อังคณาพาณิช

61102010157

อาจารย์ที่ปรึกษา

อ.ดร. ศุภร คนธกัณฑ์



01 BACKGROUND & OBJECTIVES

ที่มาและความสำคัญ
จุดประสงค์

02 SCOPE

ขอบเขตของข้อมูล

03 RESULTS

ผลการดำเนินงาน

04 DISCUSSION

สรุปและอภิปรายผลการ
ดำเนินงาน

เนื่องจากปัจจุบันมีเพลงเกิดขึ้นมากมายในประเทศไทยซึ่งในแต่ละเพลงก็ย่อมมี
อารมณ์ความหมายที่ต้องการจะสื่อแตกต่างกันออกไป และหน่วยงานบริษัทต่างๆก็ต้องการ
ที่จะวิเคราะห์เพลงเหล่านั้นเพื่อที่จะได้รู้ว่า ควรจะลงทุนหรือ สร้างเพลงใหม่ในแนวไหน

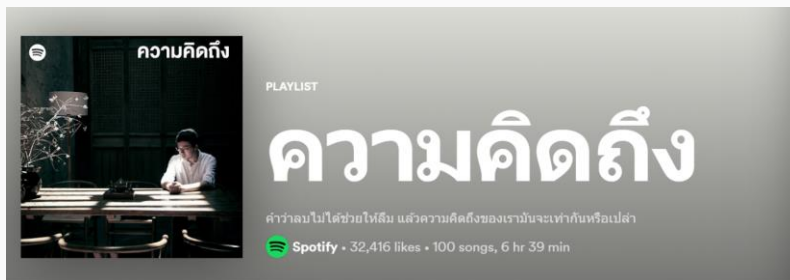
จากงานวิจัยของนิสิตปริญญาโท หลักสูตรวิทยาการข้อมูล เรื่อง “การจำแนกอารมณ์
ของเพลงไทยด้วยวิธีการเรียนรู้เชิงลึก” ที่มุ่งเน้นเรื่องการรวบรวมข้อมูลและการวิเคราะห์
เพลงไทยบนแอปพลิเคชัน “spotify” โดยใช้ข้อมูลประเภทเนื้อเพลงเป็นหลักนั้น กลุ่มของ
นิสิตเล็งเห็นว่าควรใช้ข้อมูลประเภทอื่นที่อยู่บน spotify ที่เรียกว่า “Audio feature” เข้ามา
ประกอบด้วย เพื่อช่วยในการวิเคราะห์การแบ่งอารมณ์เพลงได้อย่างมีประสิทธิภาพมาก
ยิ่งขึ้น

OBJECTIVES

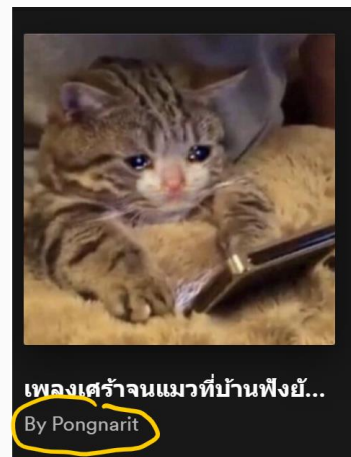
- เพื่อรวบรวมข้อมูลเพลงไทยที่อยู่ในแอปพลิเคชัน Spotify
- เพื่อศึกษาหลักการของการเรียนรู้แบบเครื่อง (machine learning) ในการวิเคราะห์ข้อมูลเพลงไทย
- เพื่อศึกษาการประยุกต์ใช้ machine learning สร้างโมเดลในการวิเคราะห์ข้อมูลเพลงไทยเพื่อจำแนกเพลงตามอารมณ์ประเภทต่างๆ ได้อย่างมีประสิทธิภาพ

SCOPE

- รวบรวมข้อมูลเพลงไทยที่อยู่ในแอปพลิเคชัน Spotify ซึ่งจะเก็บเฉพาะ playlist ที่สร้างด้วย spotify และเป็น playlist ที่สื่อถึงอารมณ์ 10 playlist



รูป 1 แสดงตัวอย่าง playlist ที่ใช้ในงานวิจัย



รูป 2 แสดงตัวอย่าง playlist ที่ไม่ใช้ในงานวิจัย

SCOPE : DATASET

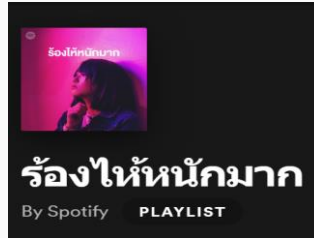
รวบรวมข้อมูลเพลงไทยที่อยู่ใน แอปพลิเคชัน Spotify

- ซึ่งอยู่บน playlist ที่สร้างด้วย spotify
- ที่เป็น playlist ที่สื่อถึงอารมณ์ 10 playlist

SCOPE : MOOD LABEL

ในงานวิจัยนี้เราจะกำหนด “อารมณ์เพลง (Mood label)” ตามชื่อ playlist ที่จัดไว้โดย spotify

CRY



SAD



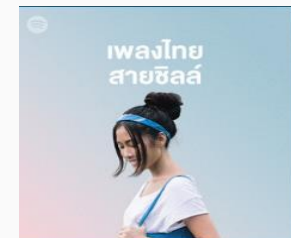
WORK



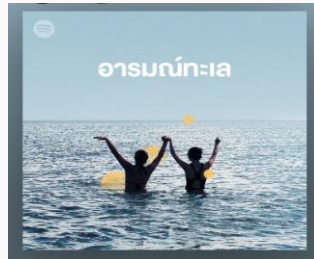
CHILLHOP



CHILLOUT



SEA



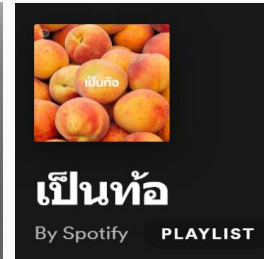
LOVESONG



TIRE

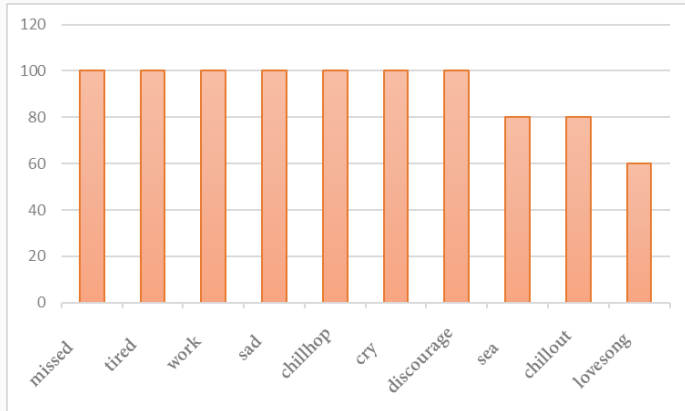


DISCOURAGE



MISSED





จำนวนเพลงทั้งหมด
= 920 เพลง



รูป 3 แสดงตัวอย่าง ข้อมูลที่ใช้ในงานวิจัย

track_name	track_id	danceability	energy	key	loudness	mode	speechiness	instrumentalness	liveness	valence	tempo	popularity	year
ปล่อย	1LOAgM7Dlxdprmoao0990N	0.697	0.356	9	-9.217	1	0.1500	1.18e-05	0.1270	0.601	66.995	53	2016
เงา	6yrxuUnWFv72Veb9jphPLL	0.523	0.615	6	-6.549	1	0.0398	0	0.1450	0.534	162.059	49	2020
แปะหัวใจ (14th Feb)	6GDI4Y8fOsptYsW2iIUbe	0.715	0.309	10	-11.412	1	0.0417	1.22e-06	0.1010	0.651	93.006	62	2021
ข้างกัน (City)	2XDCs8vAubxRe8mO1A0eTa	0.469	0.727	2	-7.484	1	0.0263	0.000918	0.2700	0.341	176.011	45	2020
สายตาหลอกกันไม่ได้ (Eyes don't lie)	4CTgPJbgnQ031q8HzhTtq2	0.552	0.796	5	-6.227	0	0.1250	1.18e-06	0.2200	0.697	172.128	60	2021
...
ฤดูเก่าในมือใหม่	7u8XLELVrKbCoGo6gBfFlo	0.699	0.406	7	-8.985	0	0.1100	0	0.2220	0.547	72.989	27	2021
ดารา	2dP91pe5MtjIzP4eYxiEt	0.833	0.431	1	-9.579	1	0.0653	3.51e-05	0.0776	0.293	112.969	26	2021
YOU KNO	37MCQy7uuQdJ41qa8m6Zh	0.739	0.690	3	-7.082	1	0.1030	0	0.1100	0.406	114.962	32	2017
จบนาว	1IGitRI6idPkjitiXsHAtN	0.760	0.637	7	-9.842	1	0.0612	0	0.1000	0.306	74.984	33	2021
LONELY	7Ca9Uubmr15yqajQFv15TR	0.761	0.547	6	-	1	0.0612	0	0.1000	0.306	74.984	33	2021
									910	0.724	80.078	30	2020

ชื่อเพลง

Audio feature

Audio feature จาก spotify

1. DANCEABILITY

ความเต้นของเพลง ให้คะแนนจาก 0.0 ไปถึง 1.0 โดยดูที่ tempo, การย่อจังหวะ, ความแข็งแรงของ beat

2. ENERGY

พลังของเพลง ให้คะแนนจาก 0.0 ถึง 1.0 โดยคำนวณจากความเร็ว ความดัง และเสียงเครื่องดนตรีบางประเภท

3.

LOUDNESS

ความดังของเพลง ซึ่งจะเป็นค่าเฉลี่ยตลอดทั้งเพลงในหน่วย decibel

4.

SPEECHINESS

เสียงพูดในเพลง จับเฉพาะเสียงที่เป็นเหมือนคนพูด (หรือ rap) ไม่ใช่เสียงร้องที่เป็นเมโลดี้

5.

- โดยทั่วไปเพลงที่ไม่มีท่อน rap จะมีค่าต่ำกว่า 0.33
- ส่วนเพลงที่มีค่าอยู่ที่ประมาณ 0.33–0.66 คือเพลงที่น่าจะมีท่อน rap ผสมอยู่
- ถ้ามากกว่า 0.66 ก็จะเป็นพวก podcast หรือ audiobook

Audio feature จาก spotify

5. LIVENESS

ตัวเลขที่แสดงว่าเพลงเป็นแนวบันทึกการแสดงสดหรือไม่

- ถ้าค่า มากกว่า 0.8 แสดงว่าเป็นเพลงที่มาจากการแสดงสด

6. VALENCE

ชั่วอารมณ์ของเพลง มีค่าตั้งแต่ 0.0 ถึง 1.0 ถ้าเพลงฟังดูแล้วสนุกสนาน ก็จะได้คะแนนสูง แต่ถ้าเป็นเพลงจังหวะช้า เศร้า ก็จะมีคะแนนน้อย

7. MODE

จะดูว่าเป็น major หรือ minor โดยดูจาก scale ของตัว melody หลัก minor แทนด้วย 0 ในขณะที่ major แทนด้วย 1

8. INSTRUMENTALNESS

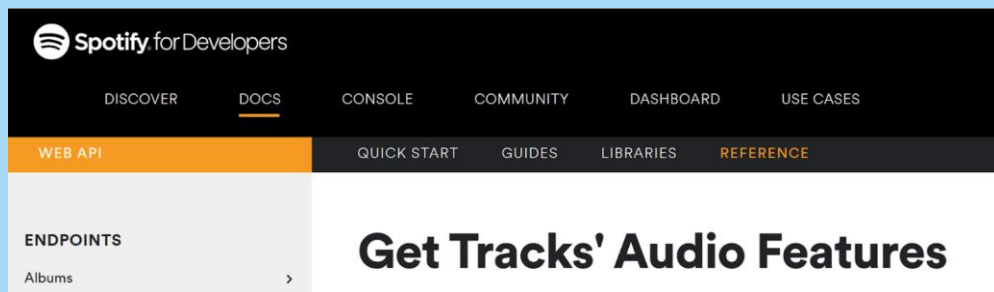
คือความน่าจะเป็นที่เพลงนี้เป็นเพลงบรรเลง โดยดูจากการการใช้เสียง โอ้วว (Ooh) หรือ อ้าา (Aah)

- ค่าใกล้ 1.0 แสดงว่าเป็นเพลงบรรเลง ไม่ค่อยมีการร้องเพลง

AUDIO FEATURE จาก SPOTIFY

9. Tempo ซึ่งเป็นค่าเฉลี่ย จังหวะเร็วช้าของเพลง ตลอดทั้งเพลงในหน่วย หน่วยเป็น bpm

10. Popularity ความนิยมของเพลง โดยคำนวณจากจำนวนครั้งที่มีการกดเล่นเพลง ในช่วงเวลาปัจจุบัน



หมายเหตุ สามารถอ่านรายละเอียดเพิ่มเติมเกี่ยวกับ Audio feature ได้ที่ [3,4]

3. RESULTS



1. ผลจากการทำ classification

จากการสังเกตจาก confusion matrix และค่า performance ของการทำ classification ด้วย โมเดลที่กำหนดไว้พบว่าเพลงทั้งหมดจาก 10 playlist หรือ label นั้น โมเดลสามารถแบ่งแยกได้ดีเพียง 3 label หรือ 4 label เท่านั้น

เราจึงตั้งสมมติฐานว่าแท้จริงแล้วในเชิง audio data นั้น เพลงจาก 10 label นั้นจะมีเพียงแค่ 3 label เท่านั้น

ตารางแสดงค่าเฉลี่ยของแต่ละโมเดล (จำนวนตัวอย่าง $N = 22$)

	Logistic regression	RandomForest	Decision Trees	Xgboost
Label =3	0.59	0.63	0.54	0.62
Label =4	0.56	0.60	0.46	0.57
Label =5	0.45	0.47	0.36	0.43

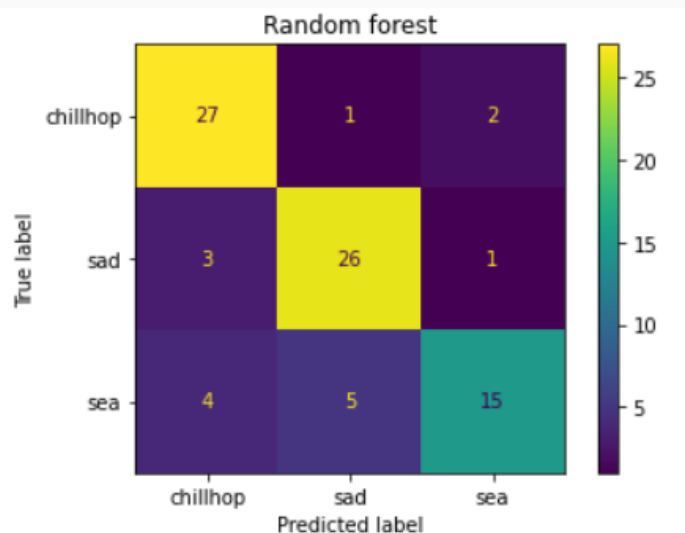
เราพบว่า **random forest** และ 3 label ให้ค่า accuracy ที่มากที่สุด

3 Label

CHILLHOP

SAD

SEA

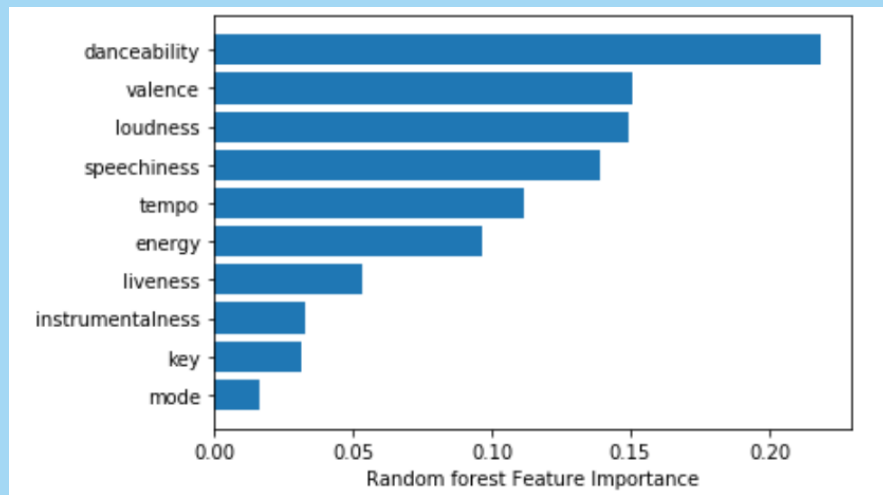


	precision	recall	f1-score	support
chillhop	0.79	0.90	0.84	30
sad	0.81	0.87	0.84	30
sea	0.83	0.62	0.71	24
accuracy			0.81	84
macro avg	0.81	0.80	0.80	84
weighted avg	0.81	0.81	0.80	84

จาก confusion matrix นี้ เราจะเห็นได้ว่า ค่าaccuracy ของการใช้ label 3 อย่างนี้ค่อนข้าง เป็นที่น่าพอใจมีการคาดคะเนผิดน้อยมาก

Feature important

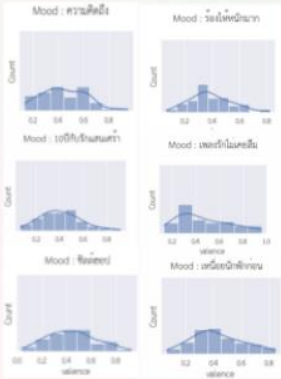
หลังจากที่เราพบว่า random forest และ 3 label ให้ค่า accuracy ที่มากที่สุด (0.81) เราได้ทำการวิเคราะห์ feature important ที่โมเดลใช้ในการตัดสินใจคัดแยกระหว่าง label พบว่า danceability valence และ loudness เป็น feature ที่โมเดลให้ความสำคัญ 3 อันดับสูงสุด



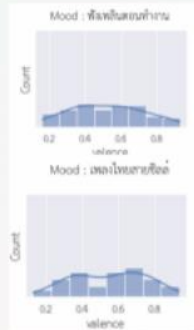
ในขั้นตอน EDA นั้นเราพบว่า Feature 3 อันดับแรก: danceability valence และ loudness นั้นมีการกระจายตัวที่สามารถแบ่งกลุ่มเพลงได้อย่างคร่าวๆ

ลักษณะการกระจายตัวของค่า valence

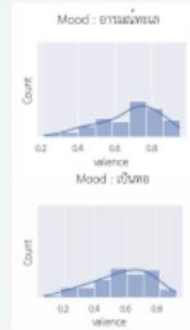
1. การกระจายตัวแบบเบ้ขวา



2. การกระจายตัวแบบ Bimodal / spread



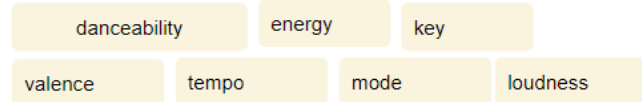
3. การกระจายตัวแบบเบ้ซ้าย



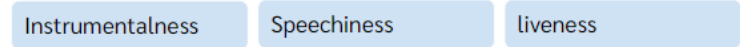
Feature selection

จากการดูการกระจายตัวของ ค่า feature ทั้ง 10 ประเภท เราพบว่า

- Feature ที่คาดว่าจะเป็น **คุณลักษณะที่สำคัญ**ในการทำ classification



- Feature ที่จะไม่นำมาใช้



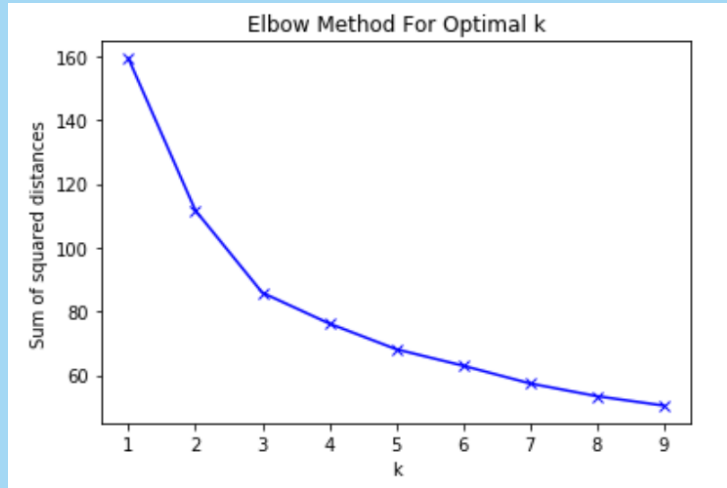
รูปแสดงตัวอย่างการทำ EDA และ Feature selection จาก presentation เมื่อวันที่ 20 ตุลาคม 2564

ผลการวิเคราะห์ จากค่า accuracy และจาก confusion matrix ที่ได้จากการทำ classification ด้วยโมเดลที่กำหนดไว้พบว่าเพลงทั้งหมดจาก 10 playlist หรือ label นั้น โมเดลสามารถแบ่งแยกได้ดีเพียง 3 label หรือ 4 label เท่านั้น

เราจึงตั้งสมมติฐานว่าแท้จริงแล้วในเชิง audio data นั้น เพลงจาก 10 label นั้นจะมี เพียงแค่ 3 label เท่านั้น

CHILLHOP	SAD	RELAXED (SEA)
	Missed Tired Feeling love (Lovesong) Cry	Chillout Focused (Work) discourage

2. ผลจากการทำ clustering โดยการใช้ K mean

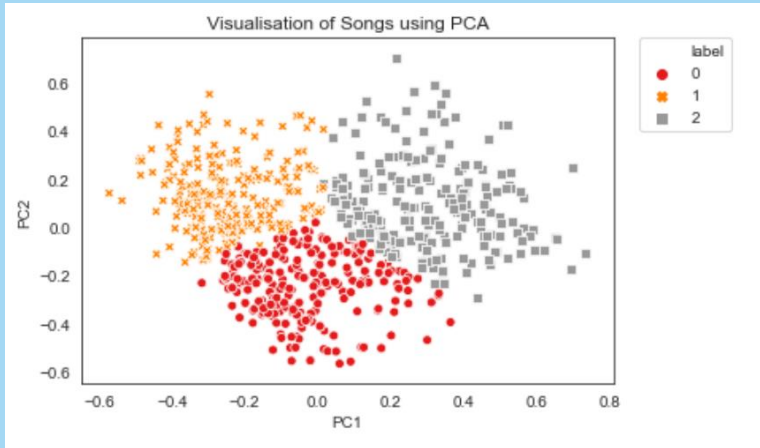


สร้าง elbow graph ซึ่งเป็นกราฟระหว่างจำนวน k (จำนวน cluster) และ ค่า sum square error

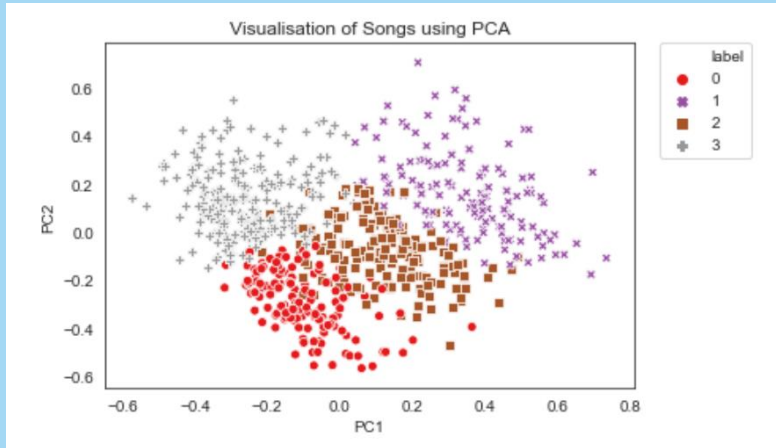
พบว่า เมื่อ $k = 3, 4$ หรือ 5 ค่า error เริ่มมีความ smooth มากขึ้น จึงทำให้เรามีความมั่นใจในสมมติฐานมากขึ้น

เพื่อให้่ายต่อการ visualize ข้อมูลมากขึ้น เราได้ทำ Dimensionality Reduction เพื่อลดมิติของข้อมูลจาก 10 มิติ เหลือเพียง 2 มิติ โดยใช้ Principal Component Analysis: PCA แล้วสังเกตการกระจายตัวของ label ที่เกิดจากการ clustering โดยใช้ K mean พบว่าข้อมูลที่อยู่ Cluster เดียวกัน มีการกระจุกตัวอยู่ใกล้ๆกัน เมื่อใช้ $k=3,4$

$k=3$

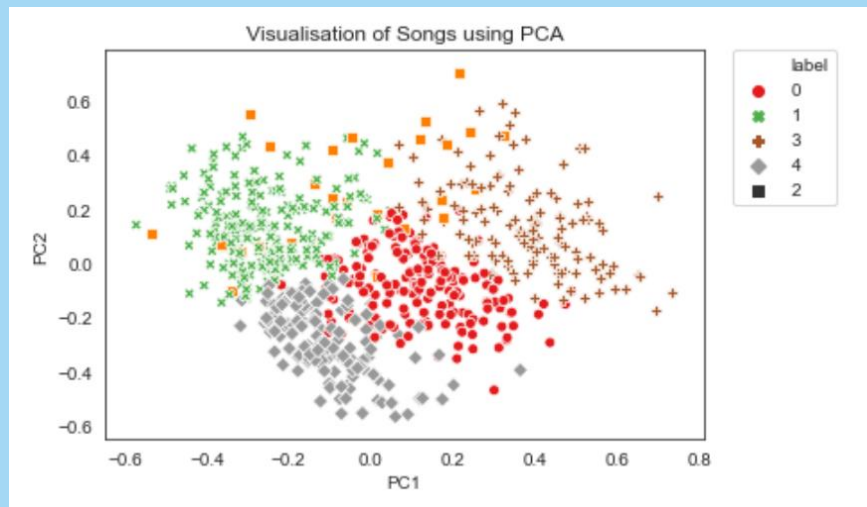


$k=4$

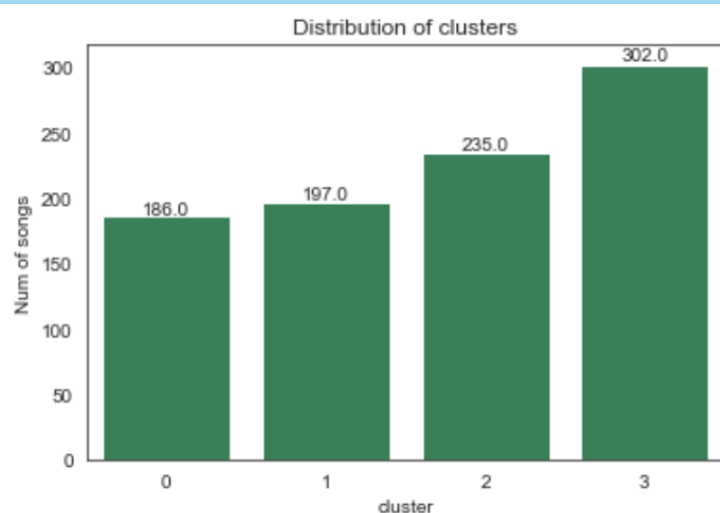
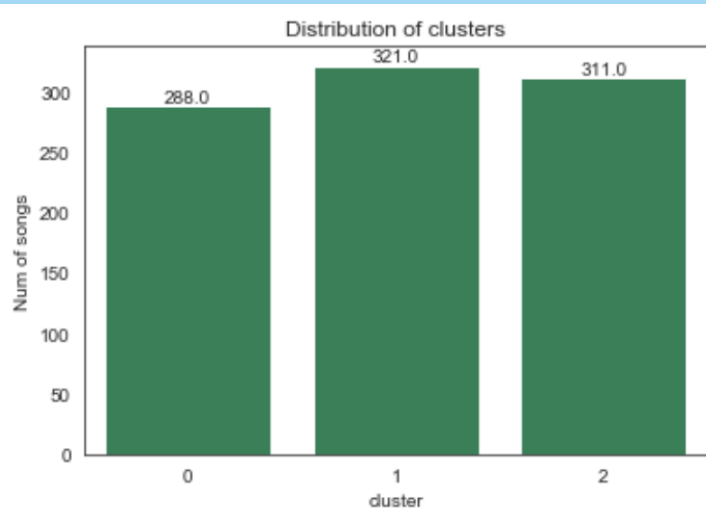


แต่เมื่อ $k=5$ พบว่า cluster ที่ 2 ข้อมูลค่อนข้างกระจายตัว และทับซ้อนกับ cluster อื่นๆ ตรงนี้ยิ่งทำให้เรามั่นใจว่าจำนวน cluster ที่ควรจะใช้แท้จริงแล้วน่าจะเป็น 3 หรือ 4 cluster เท่านั้น

$k=5$



จากนั้นเมื่อนำจำนวนข้อมูลของแต่ละ cluster มาเทียบกัน เราพบว่า เมื่อ $k=3$ จำนวนข้อมูลในแต่ละ cluster มีความใกล้เคียงกัน ในขณะที่ เมื่อ $k=4$ เราพบว่า มี 2 cluster ที่มีขนาดใกล้เคียงกัน และสังเกตว่า cluster ที่มีจำนวนข้อมูลมากที่สุด มีจำนวนข้อมูลแตกต่างจาก cluster ที่มีจำนวนน้อยที่สุดอย่างค่อนข้างมาก

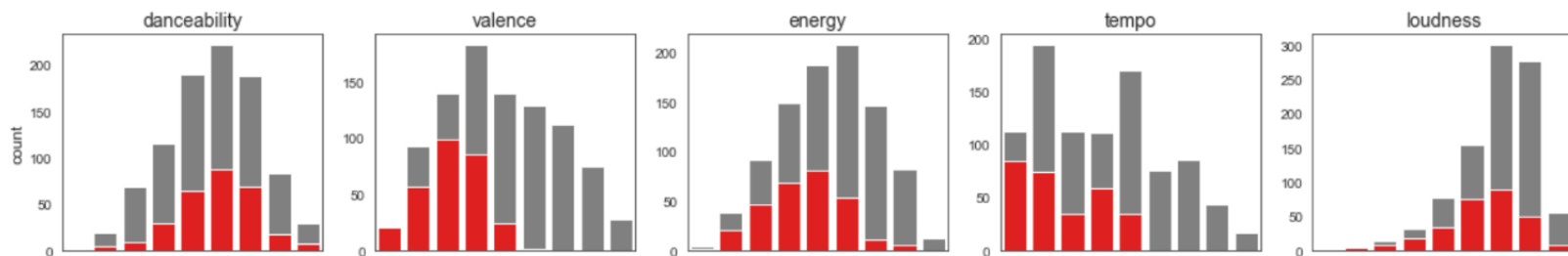


Cluster profile

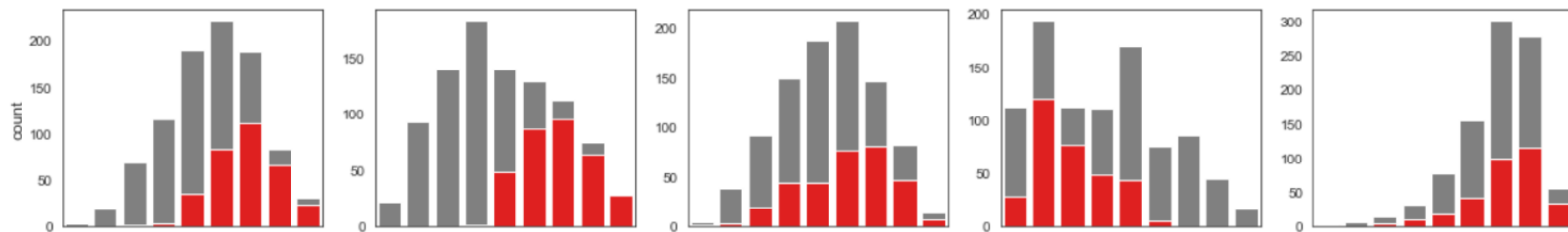
จากนั้น เราจึงได้ทำการตรวจสอบการกระจายตัวของ feature ที่มีความสำคัญ 5 อันดับแรกต่อ Randomforest พบว่าการกระจายตัวของแต่ละ cluster มีการกระจายตัวที่แตกต่างกันค่อนข้างชัดเจน โดยเฉพาะอย่างยิ่งในข้อมูล dancebility และ valence โดยที่กราฟสีเทาแสดงการแจกแจงจากข้อมูลเพลงทั้งหมด 920 เพลง กราฟสีแดงในแต่ละแถวแสดงการแจกแจงของข้อมูลเพลงในแต่ละ cluster

กราฟแสดงการกระจายตัวของแต่ละ feature ในแต่ละ cluster

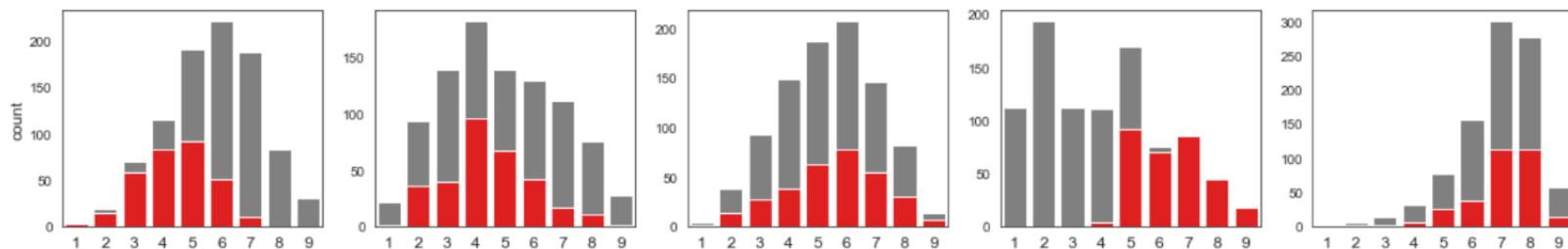
Cluster 0



Cluster 1



Cluster 2



Cluster profile

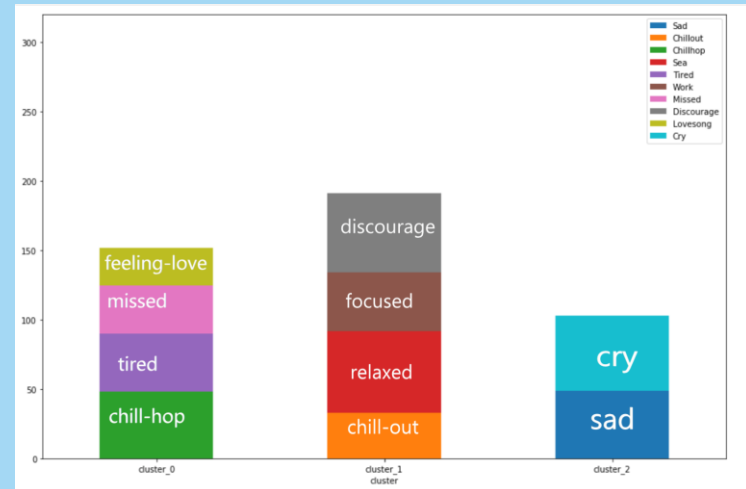
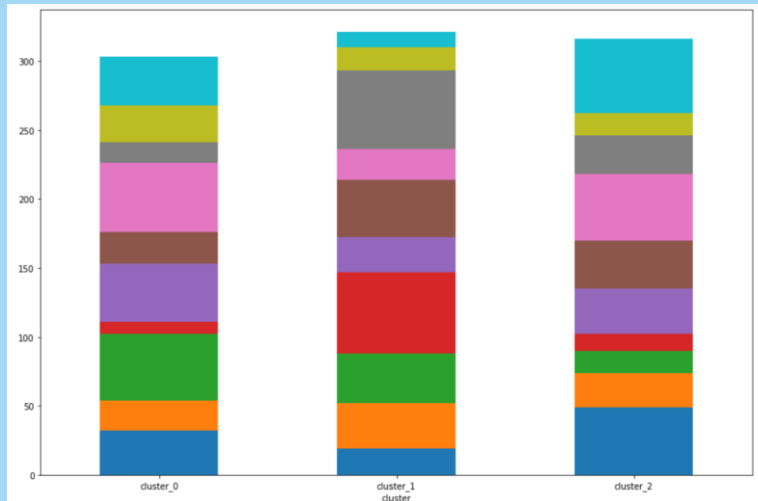
Cluster 0 : low valence, low tempo, low to average energy and average danceability
เพลงเศร้า ที่มีจังหวะช้า มีความหลากหลายของ energy และความน่าเต้น ของเพลง

Cluster 1 : high valence, low tempo , average to high energy and high danceability
เพลงมีความสุข ที่มีจังหวะช้า มีความหลากหลายของ energy และ ชวนให้รู้สึกอยากเต้น

Cluster 2 : low to average valence, high tempo, average energy and low danceability
รวมเพลงที่ให้ความเศร้า รวมไปถึง รู้สึกกลางๆ ที่มีจังหวะเร็ว มีความหลากหลายของ energy

3. อารมณ์เพลงในแต่ละ cluster

เราพบว่าแต่ละ cluster ไม่สามารถแบ่งอารมณ์ใดอารมณ์หนึ่งออกมาได้อย่างชัดเจน นั่นคือ ในอารมณ์หนึ่งๆ จะกระจายตัวไปอยู่ทุก cluster อย่างไรก็ตามเราพบว่าเมื่อเราเลือกดูเฉพาะข้อมูลส่วนใหญ่ของแต่ละอารมณ์



4. การจัดกลุ่ม label ที่ใกล้เคียงกันจากผลการ classification และ การ clustering

ตารางแสดงการจัดกลุ่มอารมณ์ทั้ง 10 ประเภทออกเป็น 3 กลุ่มใหญ่ ตามผลการทำ classification

กลุ่มอารมณ์ 1	กลุ่มอารมณ์ 2	กลุ่มอารมณ์ 3
Chillhop	Sad Missed Tired Feeling love (Lovesong) Cry	Relaxed (Sea) Chillout Focused (Work) discourage

ตารางแสดงการจัดกลุ่มอารมณ์ทั้ง 10 ประเภทออกเป็น 3 กลุ่มใหญ่ ตามผลการทำ clustering

Cluster_1	Cluster_2	Cluster_3
Sad Cry	Missed Tired Feeling love (Lovesong) Chillhop	Relaxed (Sea) Chillout Focused (Work) discourage

จากตารางการจัดกลุ่มอารมณ์ทั้ง 10 ประเภทย่อยออกเป็น 3 กลุ่มใหญ่ข้างต้น ทำให้เราสร้าง กลุ่มเพลงขึ้นมาใหม่ โดยเลือกมาเฉพาะประเภทที่เกิดการตรงกันของการทำ classification และการทำclustering

กลุ่มที่1	กลุ่มที่2	กลุ่มที่3
Cry Sad	Missed Tired Feeling love (Lovesong)	Relaxed (Sea) Chillout Focused (Work) discourage

4. DISCUSSION & SUGGESTION



สรุปผล

การวิเคราะห์ audio feature ของ 10 playlist ที่สร้างโดย spotify ที่มีการตั้งชื่อ playlist ให้สื่อถึงอารมณ์ ที่หลากหลายนั้นพบว่า

1. ใน 10 playlist นั้นแท้จริงแล้ว เป็น playlist ที่มีความคล้ายกัน นั่นคือสื่ออารมณ์ไปในทิศทางเดียวกัน
2. การจัดกลุ่มเพลงตามคล้ายคลึงกันของข้อมูล จากวิธีทั้ง 2 นั้นพบว่า ให้ผลการจัดกลุ่มบางส่วนที่เป็นไปในทิศทางด้วยกัน
 1. จากการทำ clustering profile ทำให้เห็นถึงลักษณะเพลงในกลุ่มต่างๆ
 2. ผลการจัดกลุ่มเพลงสามารถนำไปสู่การแนะนำผู้ใช้งาน (playlist suggestion) ได้

ข้อเสนอแนะ

- ❖ ในงานวิจัยนี้เราใช้ mood label จากชื่อ playlist นอกจากนี้ยังสามารถคำนวณ mood label ได้จากการใช้ Russel's model ด้วยการหาค่า Audio feature เช่น ค่า Energy, Valance
- ❖ สามารถนำไปต่อยอดโดยการสร้าง Web application เพื่อแสดงการเปรียบเทียบในเชิง audio feature ของแต่ละ playlist พร้อมกับแนะนำ playlist ที่คล้ายคลึงกันได้

Q&A

