

# Sleep Health and Lifestyle by Classification

**Student Name** Navarat Werawinantanakun

**Affiliation** Digital Business Transformation, College of Innovation,  
Thammasart University, Thailand

## Abstract

Sleep Health and Lifestyle by Classification มีวัตถุประสงค์เพื่อสร้าง Classification Model แบ่งแยกพฤติกรรมในชีวิตประจำวันที่จะส่งผลกระทบต่อพฤติกรรมการนอน เช่น อายุ อาชีพ ระดับความเครียด ค่า BMI อัตราการเต้นหัวใจ เป็นต้น จนทำให้เกิดปัญหาการนอน จาก data set ที่ประกอบไปด้วยข้อมูลจากกลุ่มตัวอย่างประมาณ 374 ตัวอย่าง

โดยผู้จัดทำได้นำ data set ผ่านกระบวนการ data exploration เพื่อดูภาพรวมและลักษณะของข้อมูล ก่อนจะทำ data pre-processing เพื่อจัดการ Missing Value, Duplicate Data และ Outlier ก่อนนำเข้าสู่ Model โดยผู้จัดทำเลือกใช้ Model K-NN และ Classification ในการประมวลผล data set นี้

โดยผลลัพธ์ที่ได้จากการทำ Classification Model ในครั้งนี้ มีประสิทธิภาพแม่นยำถึง 95% โดยสามารถคาดเดาได้ว่า จากพฤติกรรมในชีวิตประจำวันของคน ๆ หนึ่ง จะมีปัญหาในการนอนไหม เช่น หายใจขณะหลับ หรือโรคนอนไม่หลับ

**Keywords** : Classification, ปัญหาการนอน, Data Science

## Introduction

ด้วยวิถีชีวิตที่เปลี่ยนไปในปัจจุบัน ไม่ว่าจะจากหน้าที่การงาน อาหาร และการเดินทางที่ฝืดเคืองในแต่ละวัน ทำให้ร่างกายเริ่มส่งสัญญาณเตือนบางอย่างโดยที่ไม่รู้ตัว โดยเฉพาะเหล่ามนุษย์เงินเดือน และชาวออฟฟิศทั้งหลาย บางคนนำงานกลับมาทำที่บ้าน โหมงานหนักมากไป จนความเครียดสะสม กินเวลาพักผ่อนไปเรื่อย ๆ จนเริ่มมีปัญหาการนอน ซึ่งอาจส่งผลกระทบต่อการทำงานและความสัมพันธ์กับผู้อื่นได้

ตามที่ นพ.ภาณุวัฒน์ ว่องตระกูลเรือง อายุรศาสตร์ผู้สูงอายุ คลินิกผู้สูงอายุ โรงพยาบาลนครชน เผยว่า ปัญหาโรคนอนไม่หลับ หรือนอนหลับไม่สนิท สามารถเกิดได้หลายสาเหตุ ได้แก่

- ปัญหาด้านร่างกาย เช่น โรคกรดไหลย้อน, โรคสมองเสื่อมภาวะหยุดหายใจขณะหลับ
- ปัจจัยด้านจิตใจ เช่น ความเครียด ความวิตกกังวล โรคซึมเศร้า
- ปัจจัยทางสิ่งแวดล้อม เช่น แสงสว่าง เสียงดัง กลิ่นเหม็น
- ปัจจัยอื่น ๆ ที่ทำให้เกิดการนอนที่ไม่ถูกสุขลักษณะ เช่น การทำงานที่ไม่เป็นเวลา รับประทานอาหารที่ย่อยยาก ออกกำลังกายใกล้เวลานอน การเล่นเกม เครื่องดื่มที่มีคาเฟอีน

นอกจากนี้ นพ.ชาญสิริ เสกสรรวิริยะ แพทย์ศูนย์โสต ศอ นาสิก โรงพยาบาลนครชน ยังเผยว่า ภาวะอุดตันทางเดินหายใจเป็นโรคแทรกซ้อน สาเหตุมาจากสุขภาพการนอนที่ไม่ดี เป็นภาวะช่องทางเดินหายใจเกิดการอุดกั้นเป็นพัก ๆ ขณะหลับ จนเกิดการอุดตันในช่องคอแบบชั่วคราว ทำให้ลมหายใจเข้า-ออกขาดหายไปชั่วขณะ และเกิดการตื่นตัวของสมองในขณะนอนหลับเป็นพัก ๆ

ในส่วนของงานวิจัย (Grandner et.al., 2015) เขียนงานวิจัยเรื่อง “Social and Behavioral Determinants of Perceived Insufficient Sleep” กล่าวว่า การพักผ่อนไม่

เพียงพอกิน 30 วันมีความสัมพันธ์กับอายุ เพศ การศึกษา รายได้ อาชีพ พฤติกรรมการใช้ชีวิตประจำวัน เช่น พฤติกรรมการกิน การออกกำลังกาย การสูบบุหรี่ ดื่มแอลกอฮอล์ จากการศึกษากลุ่มตัวอย่างพบว่า ปัญหาการนอนไม่เพียงพอมีความสัมพันธ์กับเพศ หญิงที่ว่างงาน รายได้น้อย การศึกษาไม่สูง ขาดการออกกำลังกาย กินอาหารไม่มีประโยชน์ มีพฤติกรรมสูบบุหรี่และดื่มแอลกอฮอล์

(Dae, Hyun and Chang, 2018) เขียนงานวิจัยเรื่อง “Sleep Disturbances as a Risk Factor for Stroke” กล่าวว่า ภาวะหยุดหายใจขณะหลับ การนอนไม่หลับ และการทำงานเป็นกะ นอกจากจะทำให้การฟื้นฟูหัวใจและหลอดเลือดบกพร่องแล้ว ยังทำให้ระบบหัวใจและหลอดเลือดเกิดความเครียดด้วย เพิ่มความเสี่ยงในการเกิดโรคหลอดเลือดสมองอีกด้วย

(Azarbarzin et.al.,2020) เขียนงานวิจัยเรื่อง “The Sleep Apnea-Specific Hypoxic Burden Predicts Incident Heart Failure” กล่าวว่า ภาวะหัวใจล้มเหลว มีความเชื่อมโยงกับภาวะหยุดหายใจขณะหลับ โดยผลลัพธ์จากการทดสอบพบว่า ภาวะหยุดหายใจขณะหลับ มีความสัมพันธ์กับภาวะหัวใจล้มเหลวโดยเฉพาะในเพศชาย สามารถทำนายความเสี่ยงต่อการเกิดโรคหัวใจได้

นอกจากปัญหาที่กล่าวข้างต้น จะส่งผลกระทบต่อปัญหาการนอนแล้ว การศึกษาในครั้งนี้ ยังแสดงให้เห็นถึงปัจจัยอื่น ๆ ที่มีความสัมพันธ์อย่างมีนัยสำคัญต่ออาการเหล่านี้ด้วยเช่นกัน โดยผลลัพธ์ที่ได้จากการประมวลผลในครั้งนี้ จะเป็นแนวทางเบื้องต้นสำหรับผู้ที่กำลังประสบปัญหาการนอน ได้สำรวจพฤติกรรมในชีวิตประจำวัน ของตัวเอง และสามารถประเมินอาการเบื้องต้นได้ เพื่อลดความเสี่ยงในการเกิดโรคต่าง ๆ

## Objective

1. จัดทำ Classification Model เพื่อแสดงให้เห็นพฤติกรรมในชีวิตประจำวัน ที่ส่งผลกระทบต่อพฤติกรรมกรนอน

2. ผู้ที่สนใจหรือผู้ที่กำลังประสบปัญหาในการนอน ได้ประเมินพฤติกรรมเบื้องต้น หากเข้าข่าย จะได้ปรับปรุงพฤติกรรมเบื้องต้นได้

โดยผู้จัดทำเลือกใช้ Supervised Machine Learning อย่าง K-NN และ Decision Tree เพื่อแสดงให้เห็นพฤติกรรมต่าง ๆ ที่ส่งผลกระทบต่อพฤติกรรมกรนอน โดยจะแบ่งปัญหาการนอนออกเป็น 3 ประเภท ได้แก่

- None : ไม่มีอาการผิดปกติ
- โรคนอนไม่หลับ (Insomnia) : คนที่ประสบปัญหาในการนอนหรือนอนไม่หลับ ทำให้พักผ่อนไม่เพียงพอหรือมีคุณภาพต่ำ
- ภาวะหยุดหายใจขณะหลับ (Sleep Apnea) : ส่งผลให้การนอนหยุดชะงัก และเพิ่มความเสี่ยงทางด้านสุขภาพ

หลังจากการประมวลผล User จะสามารถประเมินพฤติกรรมเบื้องต้นได้ หากผลประเมินออกมาเข้าข่ายปัญหาดังข้างต้น User จะได้ปรับพฤติกรรมเบื้องต้นได้

## Literature Review

(Wongsirichot and Hanskunatai, 2017) เขียนงานวิจัยเรื่อง **“A Classification of Sleep Disorders with Optimal Features Using Machine Learning Techniques”** เพื่อเสนอวิธีจำแนกความผิดปกติของการนอน นอกจากวิธี PSG ซึ่งเป็นวิธีที่มีค่าใช้จ่ายสูง โดยเสนอเทคนิค opf-kNN technique ที่นอกจากประชาชนจะสามารถเข้าถึงอุปกรณ์การแพทย์นี้ได้แล้ว ยังสะดวกสบายไม่ต้องติดเซ็นเซอร์กับตัวผู้ป่วยอีกด้วย

(Lopez et.al., 2018) เขียนงานวิจัยเรื่อง **“In-Home Sleep Apnea Severity Classification using Contact-free Load Cells and an AdaBoosted Decision Tree Algorithm”** กล่าวว่า นำเสนอวิธีที่จะวินิจฉัยและจัดกลุ่มความรุนแรงของภาวะหยุดหายใจขณะหลับ โดยใช้ Algorithm Decision Tree Classifier สามารถระบุผู้ป่วยที่มีภาวะหยุดหายใจขณะหลับ และ Subsequence Linear Regression Model ประเมินการค่า AHI ปรากฏว่าประสิทธิภาพของวิธีที่งานวิจัยนี้เลือกใช้มีความแม่นยำที่ 86.97%

(Rohan and Kumari, 2021) เขียนงานวิจัยเรื่อง **“Classification of Sleep Apneas using Decision Tree Classifier”** กล่าวว่า วิธีการวินิจฉัยโรคด้วยสัญญาณ ECG เป็นวิธีที่ยากต่อการตรวจเจอ ดังนั้น การใช้คอมพิวเตอร์ในการช่วยวินิจฉัยและแยกแยะภาวะหยุดหายใจขณะหลับจึงเป็นทางเลือกที่เหมาะสม ในการศึกษาได้ใช้ Algorithm Decision Tree Classifier ได้ประสิทธิภาพมากถึง 98.39%

(Jonsson, 2022) เขียนงานวิจัยเรื่อง **“Email classification using machine learning algorithms”** เพื่อจำแนกประเภทของอีเมล และจัดหมวดหมู่อีเมลได้ดีขึ้นในอนาคต โดยใช้ Artificial neural network, K-NN, adaptive boosting และ random forest โดยผลลัพธ์ที่ได้พบว่า Artificial neural network เป็นวิธีที่เหมาะสมที่สุด

(Salari et.al., 2022) เขียนงานวิจัยเรื่อง “**Detection of sleep apnea using Machine learning algorithms based on ECG Signals: A comprehensive systematic review**” กล่าวว่า ECG เป็นวิธีที่ใช้ตรวจหาภาวะหยุดหายใจขณะหลับ ซึ่งทางผู้เขียนวิจัยต้องการหาวิธีที่สามารถตรวจหาตรวจหาภาวะหยุดหายใจขณะหลับได้ดีกว่าวิธี ECG โดยใช้ Algorithm SVM, K-NN, R-NN และพบว่า R-NN ให้ประสิทธิภาพสูงสุดถึง 99%

## ภาพรวมของ Data set

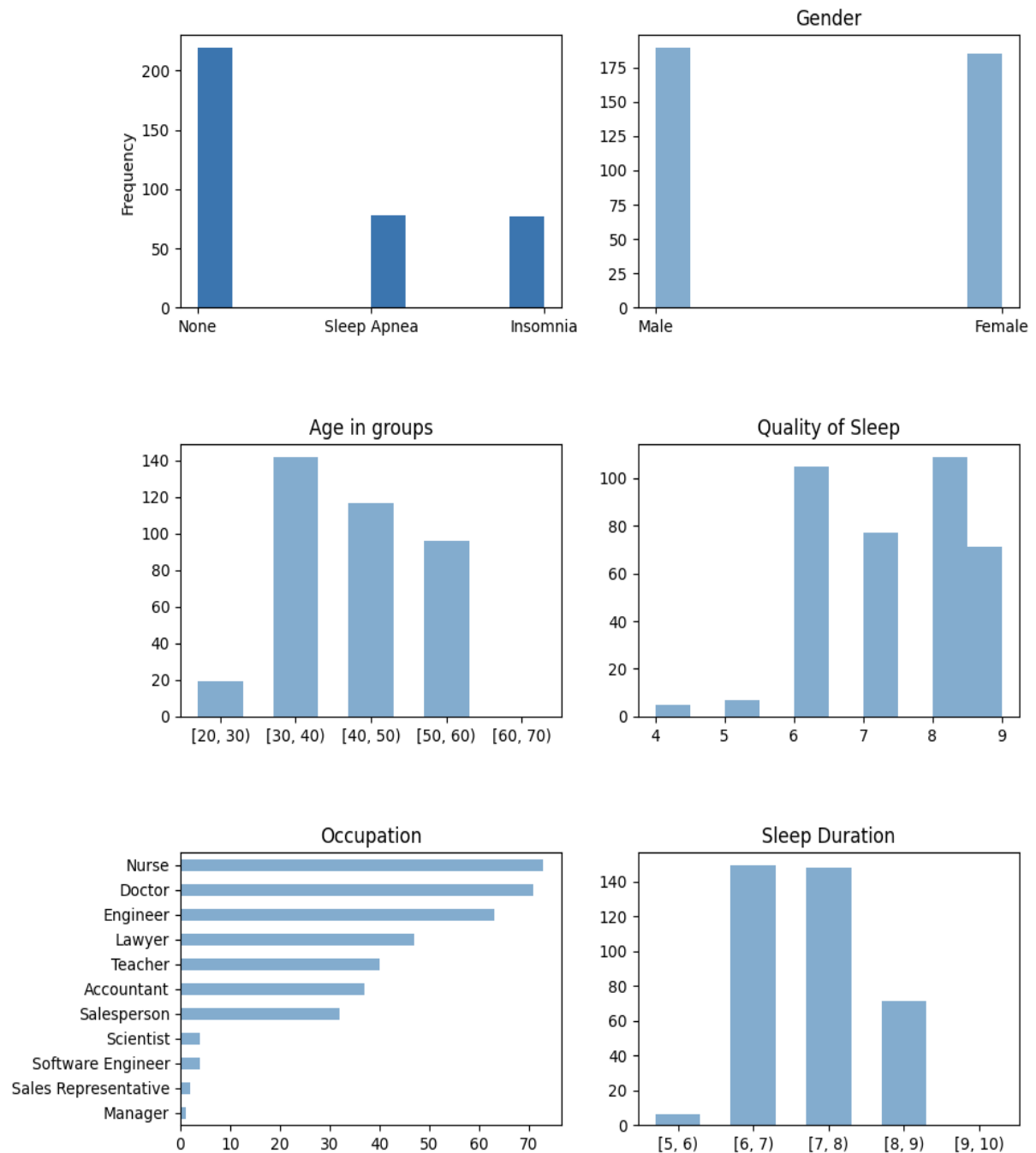
	Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
0	1	Male	27	Software Engineer	6.1	6	42	6	Overweight	126/83	77	4200	None
1	2	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None
2	3	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None
3	4	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
4	5	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
...	...	...	...	...	...	...	...	...	...	...	...	...	...
369	370	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
370	371	Female	59	Nurse	8.0	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
371	372	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
372	373	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
373	374	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea

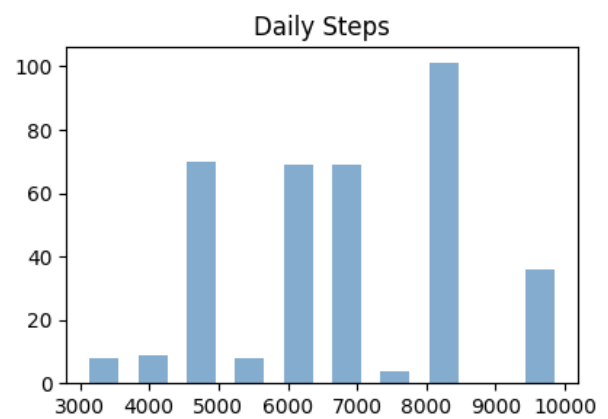
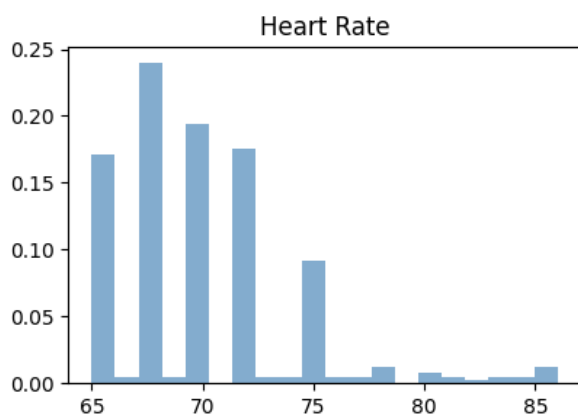
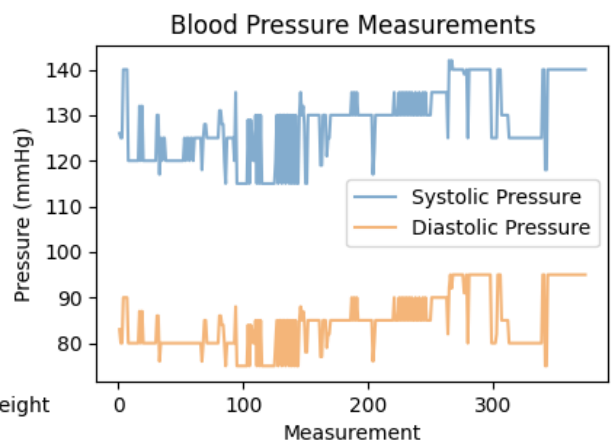
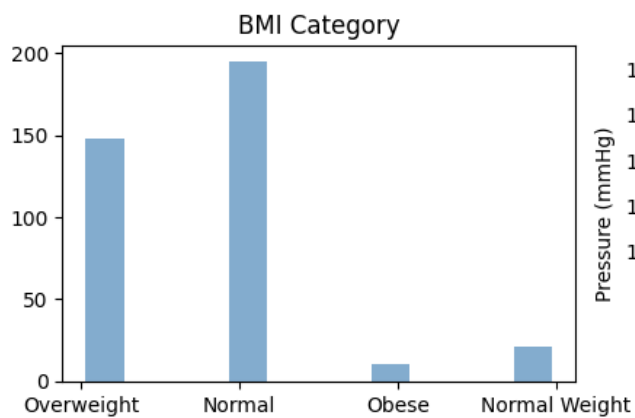
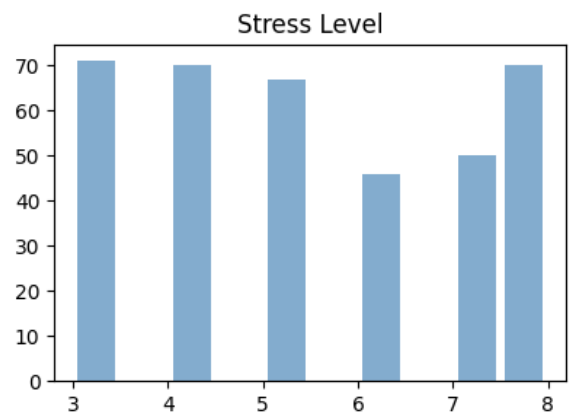
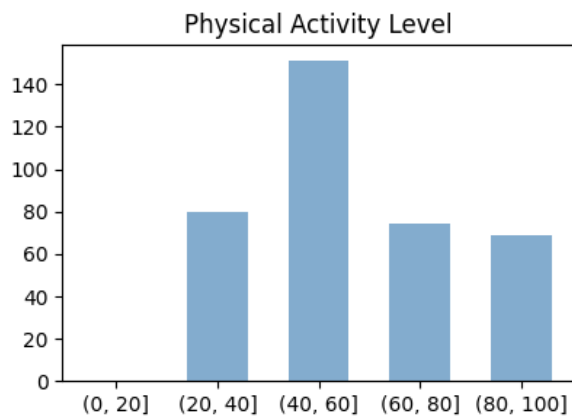
374 rows x 13 columns

Data set คำนวณโหลดมาจากเว็บไซต์ Kaggle เป็นข้อมูลเกี่ยวกับความผิดปกติในการนอน และพฤติกรรมในชีวิตประจำวันของกลุ่มตัวอย่าง ซึ่งรวมถึงข้อมูลสุขภาพของแต่ละคน เช่น อาชีพ, คุณภาพการนอน, ระดับความเครียด, ประเภทของ BMI, ความดันโลหิต, อัตราการเต้นของหัวใจ, จำนวนก้าวในแต่ละวัน และความผิดปกติในการนอน โดยคำนวณโหลดมาจากเว็บไซต์ Kaggle

ซึ่งผู้จัดทำเลือก Attribute Sleep Disorder เป็น Label เนื่องจาก ต้องการ  
สร้าง Classification Model ในการประเมินปัญหาการนอนจากพฤติกรรมใน  
ชีวิตประจำวัน

## Data Exploration

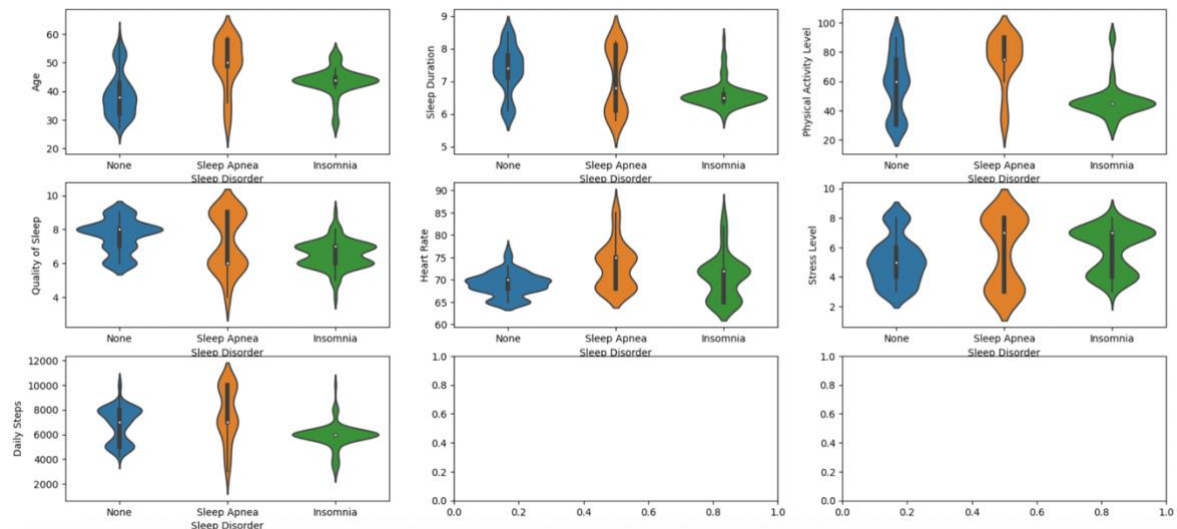




จาก Chart แสดงให้เห็นว่าแต่ละข้อมูลภายใน Attribute นั้น ๆ มีข้อมูลอะไรบ้าง และมีจำนวนเท่าไร



```
fig, ax = plt.subplots(nrows=3,ncols=3,figsize=(20,9))
ax=ax.ravel()
for i, col in enumerate(col_num):
    sns.violinplot(x='Sleep Disorder', y=col, data=df, ax=ax[i])
```



ทั้งนี้ ผู้จัดทำได้ plot chart เป็น violin โดยกำหนดให้ Label = Sleep Disorder พล็อตกับ Attribute อื่น ๆ ที่เป็นตัวเลข แสดงให้เห็นถึงความถี่ของข้อมูลว่า ปรากฏในช่วงไหนมากที่สุดในแต่ละ Attribute

## Data pre-processing

ผู้จัดทำเริ่มจากการนำ data set มาเข้ากระบวนการ data pre-processing โดยเริ่มจากการดูข้อมูลโดยรวมด้วยการเรียกใช้คำสั่ง.info() ซึ่งพบว่าในแต่ละ attribute ไม่มีค่า Missing Value อยู่

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 374 entries, 0 to 373
Data columns (total 13 columns):
#   Column                                  Non-Null Count  Dtype  
---  -
0   Person ID                               374 non-null    int64  
1   Gender                                  374 non-null    object  
2   Age                                      374 non-null    int64  
3   Occupation                              374 non-null    object  
4   Sleep Duration                          374 non-null    float64 
5   Quality of Sleep                        374 non-null    int64  
6   Physical Activity Level                 374 non-null    int64  
7   Stress Level                           374 non-null    int64  
8   BMI Category                           374 non-null    object  
9   Blood Pressure                          374 non-null    object  
10  Heart Rate                              374 non-null    int64  
11  Daily Steps                             374 non-null    int64  
12  Sleep Disorder                          374 non-null    object  
dtypes: float64(1), int64(7), object(5)
memory usage: 38.1+ KB
```

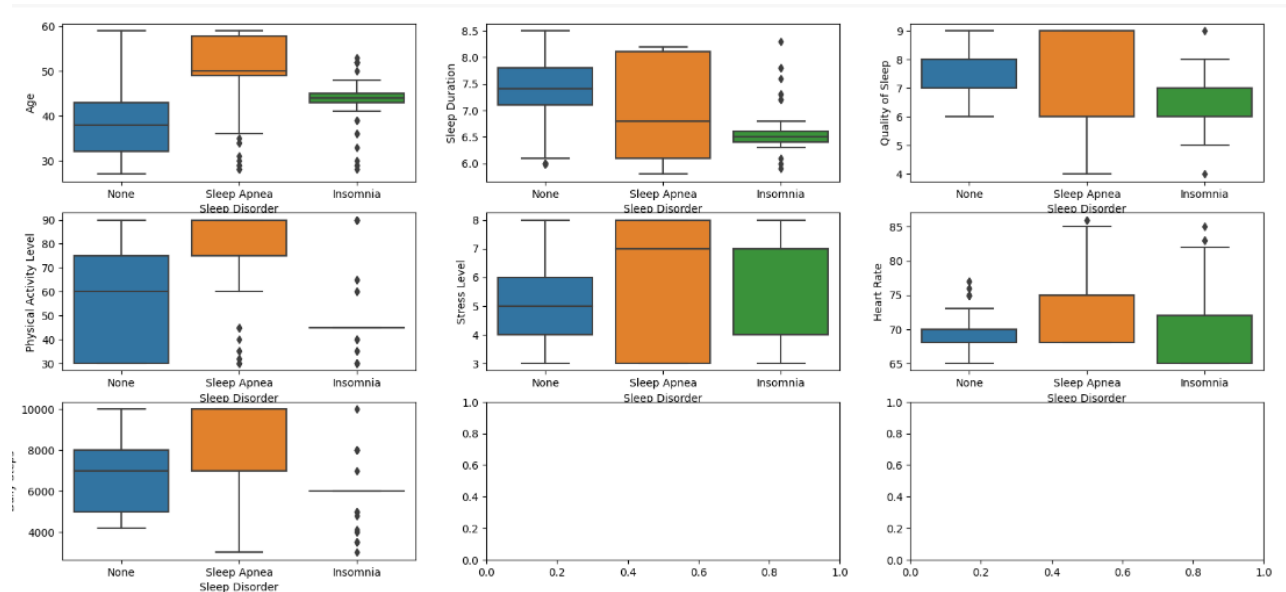
ในส่วนข้อมูล Duplicate ผู้จัดทำได้เขียน code python เพื่อ detect และ remove duplicate data ซึ่งปรากฏว่า ภายใน data set นี้ ไม่มีข้อมูลที่เป็น duplicate เนื่องจาก จำนวน row เท่าเดิม

```
df.duplicated(keep=False)
df.drop_duplicates(inplace=True)
df
```

	Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
0	1	Male	27	Software Engineer	6.1	6	42	6	Overweight	126/83	77	4200	None
1	2	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None
2	3	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None
3	4	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
4	5	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
...	...	...	...	...	...	...	...	...	...	...	...	...	...
369	370	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
370	371	Female	59	Nurse	8.0	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
371	372	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
372	373	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
373	374	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea

374 rows x 13 columns

ส่วนขั้นตอน Detect Outlier ผู้จัดทำเลือกใช้วิธี Boxplot ในการแสดงผล เพื่อเช็ค Outlier ซึ่งพบว่าบาง Attribute มีข้อมูลที่เข้าข่ายเป็น Outlier เช่น Age, Sleep Duration, Quality of Sleep, Physical Activity Level, Stress Level, Heart Rate และ Daily Steps

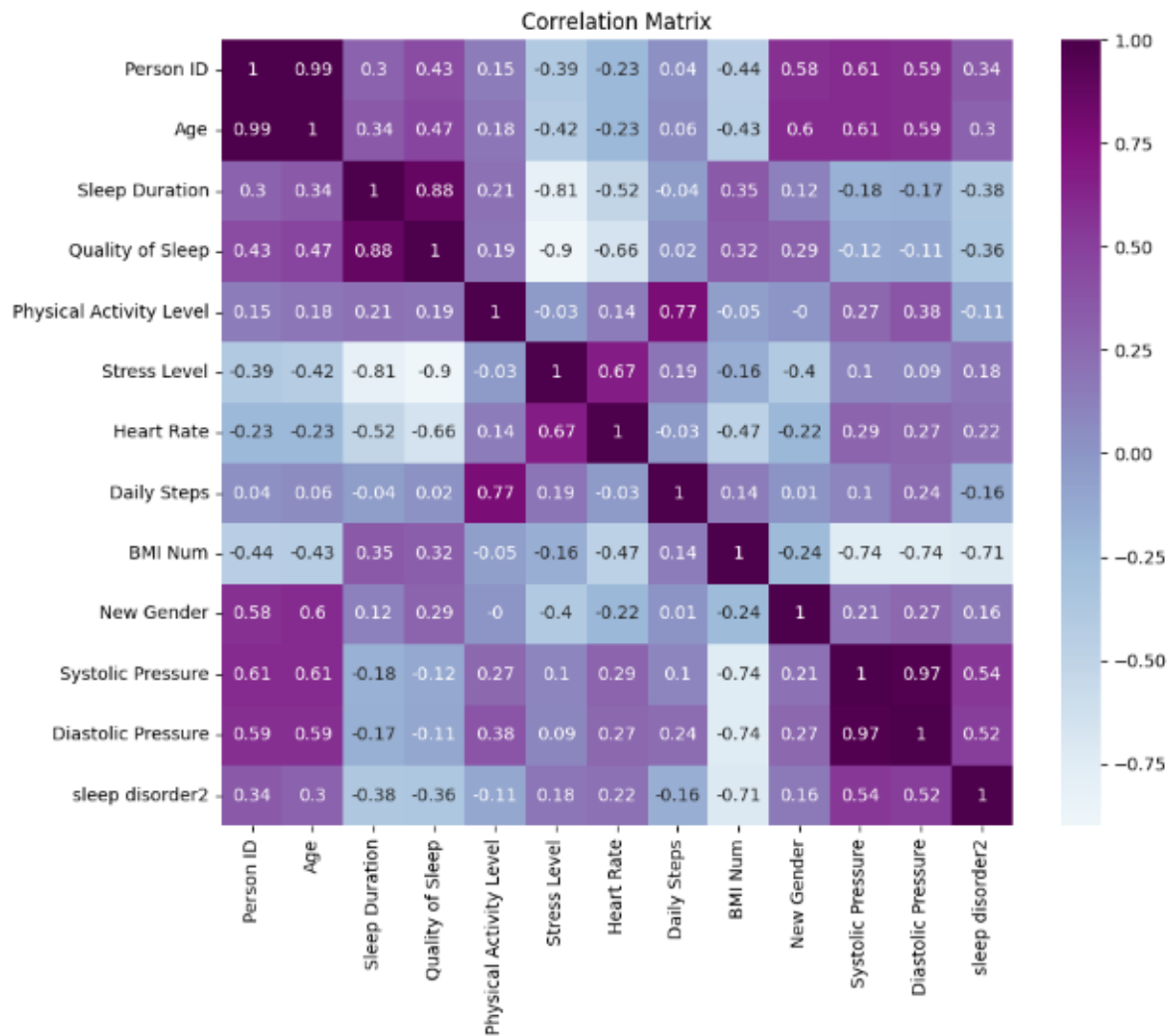


แต่ผู้จัดตัดสินใจไม่กำจัดข้อมูลเหล่านี้เนื่องจาก data set นี้มีข้อมูลค่อนข้างน้อย

นอกจากนี้ ผู้จัดทำยังได้แปลงข้อมูลจากตัวอักษรเป็นตัวเลข เพื่อความสะดวกในการทำ Machine Learning

	Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder	BMI Num	New Gender	Systolic Pressure	Diastolic Pressure	sleep disorder2
0	1	Male	27	Software Engineer	6.1	6	42	6	Overweight	126/83	77	4200	None	2	1	126	83	1
1	2	Male	28	Doctor	6.2	6	60	8	Normal Weight	125/80	75	10000	None	3	1	125	80	1
2	3	Male	28	Doctor	6.2	6	60	8	Normal Weight	125/80	75	10000	None	3	1	125	80	1
3	4	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea	1	1	140	90	2
4	5	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea	1	1	140	90	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
369	370	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea	2	2	140	95	2
370	371	Female	59	Nurse	8.0	9	75	3	Overweight	140/95	68	7000	Sleep Apnea	2	2	140	95	2
371	372	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea	2	2	140	95	2
372	373	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea	2	2	140	95	2
373	374	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea	2	2	140	95	2

และทำ Feature Selection เพื่อเลือก Attribute ที่มีสัมพันธ์กันมากที่สุดไปประมวลผล และตัด Attribute ที่แทบจะไม่มีความสัมพันธ์กันออก เพื่อให้ได้ผลลัพธ์ที่ดีที่สุด ซึ่งได้จากการทำ Correlation Matrix

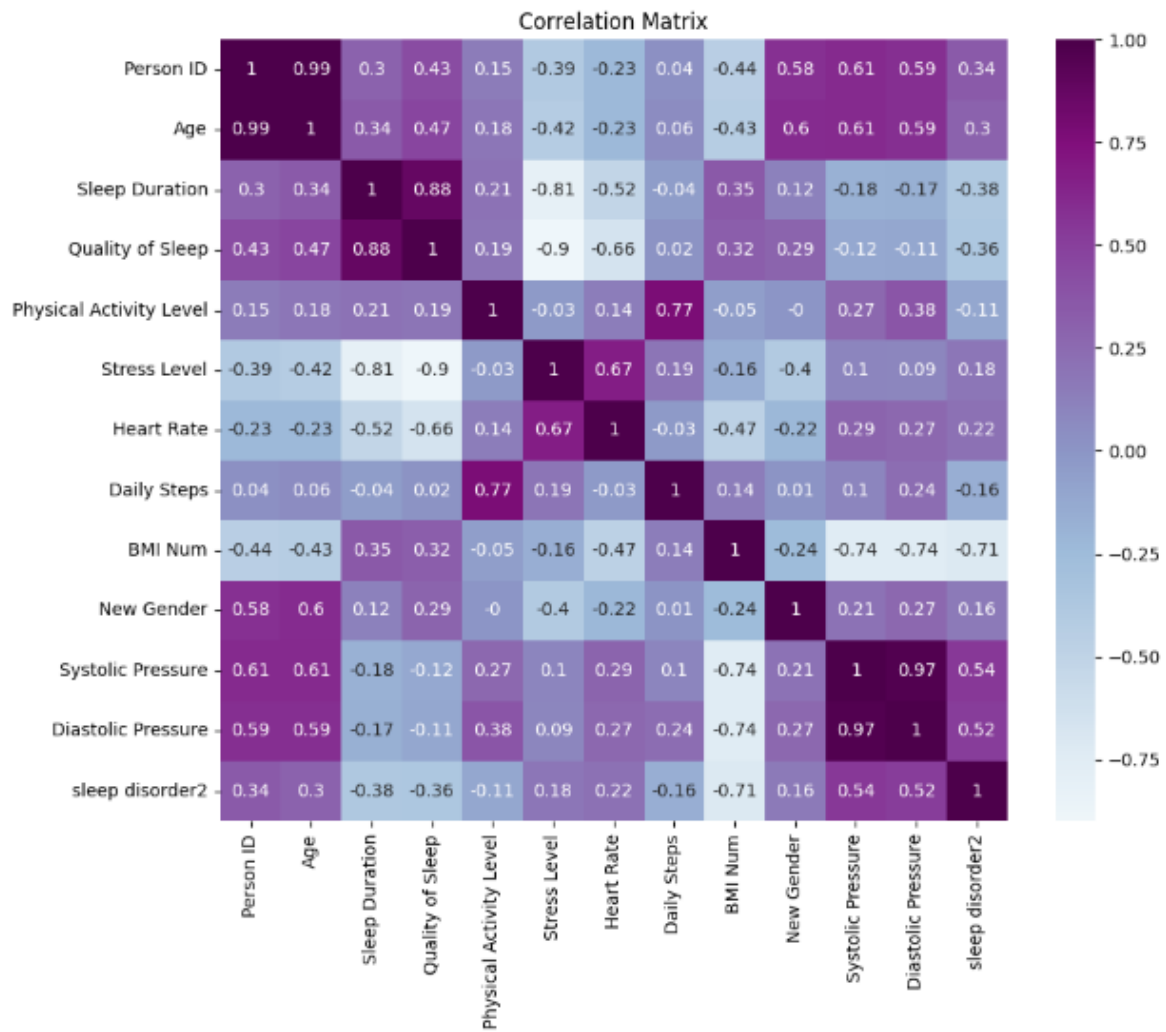


ซึ่งจะสังเกตได้ว่า เมื่อใช้ Sleep Disorder2 (Label) เป็นตัวอ้างอิงกับ Attribute อื่น ๆ พบว่า Attribute ที่มีความสัมพันธ์น้อยที่สุดที่ควรจะต้องตัดออก คือ Physical Activity, Stress Level, Daily Steps และ New Gender

## Data mining processing

เนื่องจากผู้จัดทำเลือกทำ Model แบบ Classification จึงลองทำผ่าน Algorithm ทั้งหมด 2 ตัว คือ K-NN และ Decision Tree โดยมีรายละเอียด ดังต่อไปนี้

### 1. K-NN



- จาก Correlation Matrix ผู้จัดทำได้ลองตัด Attribute ที่มีความสัมพันธ์น้อยแล้วพบว่า การตัด Physical Activity Level, Stress Level, Daily Steps และ New Gender ทำให้ประสิทธิภาพของ K-NN Model มีประสิทธิภาพดีที่สุด

```
from sklearn.datasets import make_blobs
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
```

```
x = df[column_num]
y = df['Sleep Disorder']
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.2)
scaler = StandardScaler()
x_train = scaler.fit_transform(x_train)
x_test = scaler.fit_transform(x_test)
```

```
knn = KNeighborsClassifier(n_neighbors=10)
knn.fit(x_train, y_train)
```

```
▼ KNeighborsClassifier
KNeighborsClassifier(n_neighbors=10)
```

ble-click (or enter) to edit

```
y_pred = knn.predict(x_test)
y_pred
from sklearn.metrics import classification_report, confusion_matrix
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

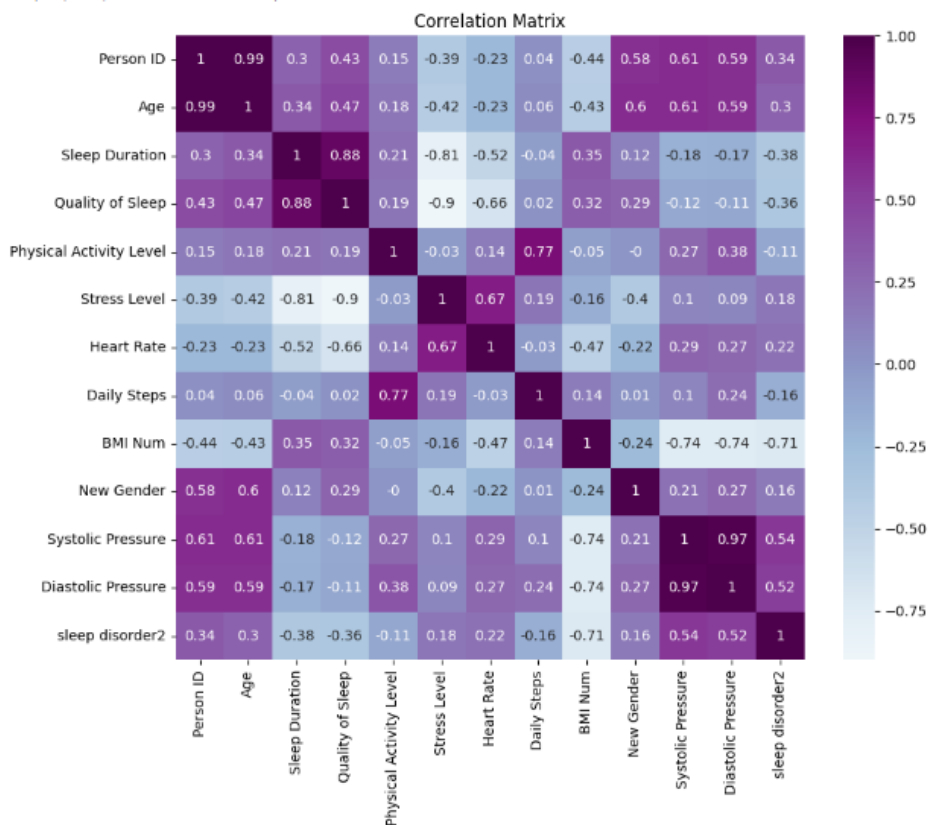
```
[[11  1  1]
 [ 1 42  1]
 [ 0  1 17]]
```

	precision	recall	f1-score	support
Insomnia	0.92	0.85	0.88	13
None	0.95	0.95	0.95	44
Sleep Apnea	0.89	0.94	0.92	18
accuracy			0.93	75
macro avg	0.92	0.92	0.92	75
weighted avg	0.93	0.93	0.93	75

- Import make\_blobs เข้ามาจาก sklearn.datasets
- Import KNeighborsClassifier เข้ามาผ่าน sklearn.neighbors
- Import train\_test\_split เข้ามาจาก sklearn.model\_selection เพื่อแบ่งข้อมูลเป็น train set และ test set
- Import modul StandardScaler เพื่อปรับ scale ข้อมูลให้อยู่ในระดับใกล้เคียงกัน

- โดยกำหนดให้ x เก็บข้อมูล feature อื่น ๆ และให้ y เก็บข้อมูล Label (Sleep Disorder)
- กำหนด Train set = 80% และให้ Test set = 20%
- กำหนด n\_neighbors = 10 เนื่องจาก ให้ผลลัพธ์ที่ดีที่สุด
- กำหนดให้ y\_pred เก็บค่า predictที่ได้จาก x\_test
- Import classification\_report เพื่อแสดงผลประสิทธิภาพของ Model K-NN ซึ่งผลลัพธ์ที่ได้ คือ 95%

## 2. Decision Tree



- จาก Correlation Matrix ผู้จัดทำได้ลองตัด Attribute ที่มีความสัมพันธ์น้อยแล้วพบว่า การตัด Physical Activity Level ทำให้ประสิทธิภาพของ Decision Tree มีประสิทธิภาพดีที่สุด

```
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
x=df[col_tree]
y=df['Sleep Disorder']
test_size=0.2
x_train, x_test, y_train, y_test = train_test_split(x,y,
                                                    test_size=test_size,
                                                    stratify=y)

model=DecisionTreeClassifier(criterion='gini')
model.fit(x_train, y_train)
```

```
▼ DecisionTreeClassifier
DecisionTreeClassifier()
```

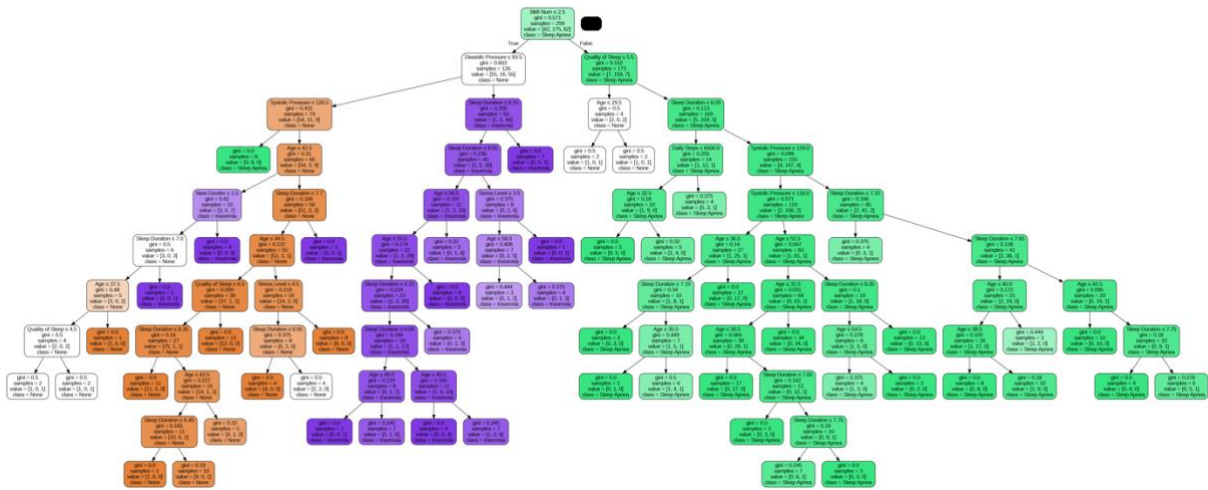
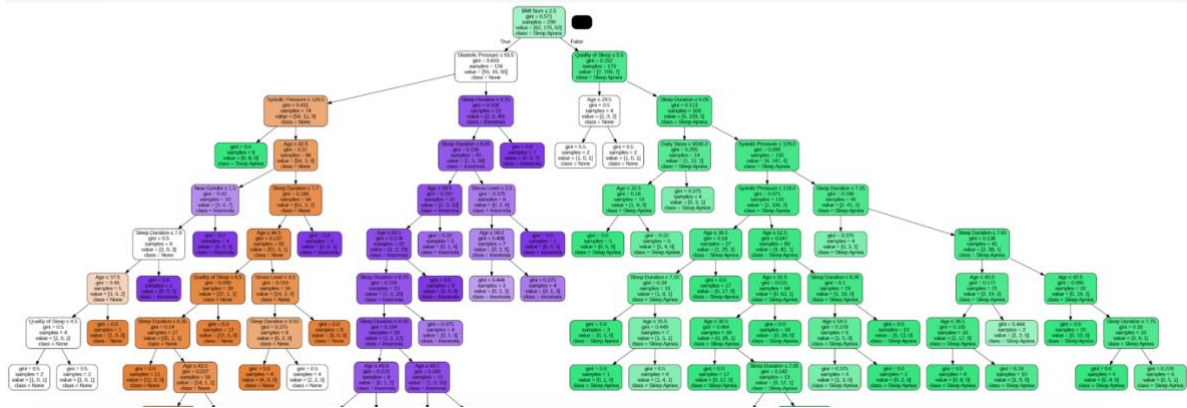
- สร้างตัวแปร col\_tree เพื่อเก็บข้อมูลทุก Attribute ยกเว้น Attribute Sleep Disorder เนื่องจาก ต้องการกำหนดให้เป็น label
- Import train\_test\_split จาก sklearn.model\_selection เพื่อข้อมูลเป็น train set และ test set
- Import DecisionTreeClassifier จาก sklearn.tree
- กำหนด x เก็บค่า col\_tree จากตัวแปร sleep
- กำหนด y เก็บค่า Attribution Sleep Disorder
- กำหนด test\_size โดยกำหนดให้ข้อมูล train set = 80% และ test set = 20%
- สร้าง Model decision tree ด้วยคำสั่ง DecisionTreeClassifier



```

from io import StringIO
from sklearn.tree import export_graphviz
import pydotplus
from IPython.display import Image
dot_data = StringIO()
export_graphviz(model, out_file=dot_data,
               feature_names=feature_names,
               class_names=class_names,
               filled=True, rounded=True,
               special_characters=False)
graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
Image(graph.create_png())

```



- สร้าง visual ของ Decision Tree ออกมาจะได้ผลลัพธ์ตามภาพข้างต้น
- กำหนดให้ predict เก็บค่า predictที่ได้จาก x\_test

```

predicted= model.predict(x_test)
predicted

array(['None', 'Insomnia', 'None', 'None', 'None', 'None', 'Sleep Apnea',
       'None', 'None', 'None', 'None', 'None', 'Insomnia', 'Insomnia',
       'None', 'Insomnia', 'None', 'Sleep Apnea', 'None', 'None',
       'Insomnia', 'Sleep Apnea', 'Insomnia', 'None', 'Insomnia', 'None',
       'None', 'None', 'Insomnia', 'Sleep Apnea', 'None', 'Insomnia',
       'Sleep Apnea', 'Insomnia', 'None', 'None', 'None', 'Insomnia',
       'Sleep Apnea', 'Sleep Apnea', 'None', 'None', 'None', 'Insomnia',
       'None', 'None', 'Sleep Apnea', 'None', 'None', 'None', 'Insomnia',
       'None', 'Sleep Apnea', 'None', 'None', 'None', 'None', 'None',
       'Sleep Apnea', 'None', 'None', 'None', 'Insomnia', 'Insomnia',
       'None', 'None', 'Sleep Apnea', 'None', 'None', 'None',
       'Sleep Apnea', 'None', 'Sleep Apnea', 'Sleep Apnea', 'Sleep Apnea'],
      dtype=object)

```

```

from sklearn.metrics import classification_report, confusion_matrix
print(confusion_matrix(y_test, predicted))
print(classification_report(y_test, predicted))

```

```

[[13  2  0]
 [ 1 43  0]
 [ 1  0 15]]

```

	precision	recall	f1-score	support
Insomnia	0.87	0.87	0.87	15
None	0.96	0.98	0.97	44
Sleep Apnea	1.00	0.94	0.97	16
accuracy			0.95	75
macro avg	0.94	0.93	0.93	75
weighted avg	0.95	0.95	0.95	75

- Import `classification_report` เพื่อแสดงผลประสิทธิภาพของ Model Decision Tree ซึ่งผลลัพธ์ที่ได้ คือ 95%

## Conslusion

**Sleep Health and Lifestyle by Classification** เป็น data set เกี่ยวกับปัญหาการนอน ประกอบกับพฤติกรรมในชีวิตประจำวัน ซึ่งผู้จัดทำนำ data set นี้ มาจัดทำ Classification ผ่าน Model 2 แบบ คือ K-NN และ Decision Tree พบว่า Model Decision Tree ให้ประสิทธิภาพดีที่สุดอยู่ที่ 95% ซึ่งสามารถนำไปใช้กับประชาชนทั่วไปในการคาดการณ์เบื้องต้นว่า เขามีปัญหาการนอนไหม

ทั้งนี้ ปัญหาในการทำโปรเจกในครั้งนี้ คือ ความรู้ด้าน Machine Learning ยังไม่หลากหลายที่จะใช้ Model อื่น ๆ ที่อาจจะให้ประสิทธิภาพดีกว่ามาทำให้ และสำหรับผู้ ที่สนใจนำ Model ของ data mining นี้ไปใช้ต่อยอดในเชิงการแพทย์ หรือทางสถิติได้

## Reference

### หนังสือและเอกสารตีพิมพ์

Azarbarzin et.al.(2020). *The Sleep Apnea-Specific Hypoxic Burden Predicts Incident Heart Failure : Dissertation*. Retrieved from National Library of Medicine.

Dae, Hyun and Chang. (2018). *Sleep Disturbances as a Risk Factor for Stroke. : Dissertation*. Retrieved from National Library of Medicine.

Grandner et.al. (2015). *Social and Behavioral Determinants of Perceived Insufficient Sleep. : Dissertation*. Retrieved from National Library of Medicine.

Jonsson, I.(2022). *Email classification using machine learning algorithms*. Retrieved from Uppsala University.

Lopez et.al. ( 2018) . *In-Home Sleep Apnea Severity Classification using Contact-free Load Cells and an AdaBoosted Decision Tree Algorithm*. Retrieved from Uppsala University.

Rohan and Kumari.(2021). *Classification of Sleep Apneas using Decision Tree Classifier*. Retrieved from researchgate.

Wongsirichot and Hanskunatai. 2017. *A Classification of Sleep Disorders with Optimal Features Using Machine Learning Techniques*. Dissertation. Retrieved from Faculty of Science, King Mongkut's Institute of Technology Ladkrabang

## ข้อมูลจากเว็บไซต์ภาษาไทย

ชาญสิริ เสกสรรค์วิริยะ. นอนกรน หายุดหายใจขณะหลับ เสี่ยงกว่าที่คิด แต่รักษาได้. เข้าถึงเมื่อ 30 กรกฎาคม 2566 เข้าถึงได้จาก

<https://www.nakornthon.com/article/detail>

ภาณุวัฒน์ ว่องตระกูลเรือง. โรคนอนไม่หลับ (insomnia) เข้าถึงเมื่อ 30 กรกฎาคม 2566 เข้าถึงได้จาก <https://www.nonthavej.co.th/Insomnia-2.php>

## Appendix

### Python Code :

[https://colab.research.google.com/drive/1jcBanXzgXv\\_46BClkELsCtP9v6W8bX3Q?usp=sharing](https://colab.research.google.com/drive/1jcBanXzgXv_46BClkELsCtP9v6W8bX3Q?usp=sharing)

Original Data Set : <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset?fbclid=IwAR2A4AweVm5TrytJLnBlq1Lt2iDGEH9y26OSCnBjry8CjdimpSP7NMp8BFE>

Example ၅၀၄ data set

	Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
0	1	Male	27	Software Engineer	6.1	6	42	6	Overweight	126/83	77	4200	None
1	2	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None
2	3	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None
3	4	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
4	5	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
...	...	...	...	...	...	...	...	...	...	...	...	...	...
369	370	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
370	371	Female	59	Nurse	8.0	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
371	372	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
372	373	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
373	374	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea

374 rows x 13 columns