

การคาดการณ์แนวโน้มความชุกและปัจจัยที่สัมพันธ์กับโรคอ้วน
ของผู้ป่วยความดันโลหิตสูงในประเทศไทย

PREDICTION OF PREVALENCE TRENDS AND ASSOCIATED FACTORS
FOR OBESITY OF HYPERTENSIVE PATIENTS IN THAILAND



นางสาวกุลยา วิสुทธิเทพ
KUNLAYA WISUTTHEP

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์
ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2565

KMITL-2022-SC-M-017-102

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

PREDICTION OF PREVALENCE TRENDS AND ASSOCIATED FACTORS
FOR OBESITY OF HYPERTENSIVE PATIENTS IN THAILAND



AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN
DATA SCIENCE AND ANALYTICS
KMITL-DIGITAL ANALYTICS AND INTELLIGENCE CENTER SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2022

KMITL-2022-SC-M-017-102

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2022

SCHOOL OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ใบรับรองการค้นคว้าอิสระ

หัวข้อการค้นคว้าอิสระ

การคาดการณ์แนวโน้มความชุกและปัจจัยที่สัมพันธ์กับโรคอ้วนของ
ผู้ป่วยความดันโลหิตสูงในประเทศไทย
Prediction of Prevalence Trends and Associated Factors for
Obesity of Hypertensive Patients in Thailand

ชื่อนักศึกษา

นางสาวกัญญา วิสุทธิ์เทพ

รหัสประจำตัว

63605085




ปริญญา

วิทยาศาสตรมหาบัณฑิต (วิทยาการข้อมูลและการวิเคราะห์)

ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง

อาจารย์ที่ปรึกษาการค้นคว้าอิสระ

ผู้ช่วยศาสตราจารย์ ดร.วรางคณา กัมปาน

คณะกรรมการสอบการค้นคว้าอิสระ	ลายมือชื่อ
ดร.จิรภัทร์ หยกรัตนศักดิ์ ประธานกรรมการ	
ดร.ยุวดี กล่อมวิเศษ กรรมการ	
ผู้ช่วยศาสตราจารย์ ดร.วรางคณา กัมปาน กรรมการ	

วัน/ เดือน/ ปี ที่สอบ วันอาทิตย์ที่ 17 กรกฎาคม พ.ศ. 2565 เวลา 16.00 – 19.00 น.

สถานที่สอบ ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง

คณะวิทยาศาสตร์รับรองแล้ว



(รองศาสตราจารย์ ดร.สุธี ชูดีไพจิตร)

คณบดีคณะวิทยาศาสตร์

วันที่ 27 กรกฎาคม 2565

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อการค้นคว้าอิสระ	การคาดการณ์แนวโน้มความชุกและปัจจัยที่สัมพันธ์กับโรคอ้วนของผู้ป่วยความดันโลหิตสูงในประเทศไทย
ชื่อนักศึกษา	นางสาวกุลยา วิสุทธิ์เทพ
รหัสประจำตัว	63605085
ปริญญา	วิทยาศาสตร์มหาบัณฑิต (วิทยาการข้อมูลและการวิเคราะห์)
พ.ศ.	ศุนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง 2565
อาจารย์ที่ปรึกษาการค้นคว้าอิสระ	ผู้ช่วยศาสตราจารย์ ดร.วรางคณา กัมปาน

บทคัดย่อ

โรคความดันโลหิตสูงเป็นโรคไม่ติดต่อเรื้อรัง (Non-communicable diseases : NCDs) ชนิดหนึ่งซึ่งผลกระทบโดยรวมต่อสุขภาพของผู้ป่วยเนื่องจากโรคความดันโลหิตสูงเป็นปัจจัยที่ทำให้เกิดโรคไม่ติดต่อเรื้อรังชนิดอื่นๆอีกมากมายรวมถึงภาวะหัวใจวาย และอาจจะนำไปสู่สาเหตุการเสียชีวิตได้

จากผลการวิเคราะห์จากกลุ่มผู้ป่วยความดันโลหิตสูงในประเทศไทยในปี พ.ศ. 2557 พ.ศ. 2558 และ พ.ศ. 2561 พบว่าสัดส่วนของผู้ป่วยความดันโลหิตสูงที่มีโรคอ้วนร่วมด้วยสูงถึง 46.19% จากจำนวนตัวอย่างทั้งหมด 79,663 รายการ กระจายตัวอยู่ในทั่วทุกภูมิภาคของประเทศไทย โดยเฉพาะกรุงเทพมหานคร และมีแนวโน้มที่คาดว่าจะสูงขึ้นเช่นเดียวกัน

ดังนั้นการลดปัจจัยเสี่ยงที่จะก่อให้เกิดโรคความดันโลหิตสูงหรือบรรเทาอาการลงจึงเป็นเรื่องที่สำคัญ หนึ่งในวิธีเหล่านั้นคือการลดน้ำหนักในผู้ที่มีน้ำหนักเกินหรือดัชนีมวลกายมากกว่า 22.9 กก./ตร.กก. จึงเป็นที่มาของการจัดทำการศึกษาครั้งนี้โดยใช้ตัวแบบทางวิทยาการข้อมูลคือ Logistic Regression เข้ามาช่วยแสดงความสัมพันธ์และแนวโน้มของโรคอ้วนในผู้ป่วยความดันโลหิตสูงเพื่อสะท้อนให้เห็นถึงปัญหาและประสิทธิภาพการส่งเสริมสุขภาพโดยเริ่มจากภาพรวมภายในประเทศ เพื่อค้นหาปัจจัยที่มีความสัมพันธ์ต่อการควบคุมน้ำหนักเพื่อนำไปสู่การพัฒนานโยบายการควบคุมโรคความดันโลหิตสูงต่อไป

คำสำคัญ : การเป็นโรคอ้วน ความสัมพันธ์ของปัจจัย แนวโน้ม ภาวะน้ำหนักเกิน โรคความดันโลหิตสูง โรคไม่ติดต่อเรื้อรัง

Independent Study Title	Prediction of Prevalence Trends and Associated Factors for Obesity of Hypertensive Patients in Thailand
Student Name	Kunlaya Wisuttthep
Student ID	63605085
Degree	Master of Science (Data Science and Analytics) KMITL-Digital Analytics and Intelligence Center
Year	2022
Independent Study Advisor	Asst.Prof.Dr.Warangkhana Kimpan

Abstract

Hypertension is one of Chronic Non-Communicable Diseases (NCDs) that affects overall health of patients. It is a factor that can cause many other chronic non-communicable diseases including heart failure which may lead to the cause of death.

According to the analysis of hypertensive patients in Thailand in 2014, 2015, and 2018, the proportion of hypertensive patients with obesity accounted for 46.19% from 79,633 sample of data records around every region in Thailand. Moreover, it is likely to increase especially in Bangkok.

Therefore, reducing the risk factors of hypertension or alleviating the symptoms are very important. One of the solutions is losing weight in those who has BMI greater than 22.9 kg/m². This independent study aims to utilize the Data Science model, called Logistic Regression to show the relationship and trend of obesity in hypertensive patients to reflect problems for health promotion effectiveness-in Thailand and find associated factors affecting weight control, leading to further development of hypertension control policies.

Keywords : Associated Factors, Hypertensive, Non-communicable Diseases, Obesity, Overweight, Trends

กิติกรรมประกาศ

งานค้นคว้าอิสระชิ้นนี้สำเร็จด้วยจากความอนุเคราะห์ของหลายๆท่านโดยเฉพาะอย่างยิ่งคณาจารย์ประจำหลักสูตรวิทยาการข้อมูลและการวิเคราะห์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่ช่วยประสิทธิ์ประสาทความรู้ให้ก่อเกิดเป็นผลงานที่สามารถสร้างประโยชน์ให้แก่ผู้อื่นต่อไปได้

ขอขอบพระคุณอาจารย์ที่ปรึกษา ผศ.ดร.วรางคณา กิมปาน ที่ช่วยเหลือ ผลักดัน ตรวจสอบความถูกต้องและติดตามผลทำให้ผลงานค้นคว้าอิสระชิ้นนี้สำเร็จตรงตามเป้าหมายได้อย่างมีคุณภาพ

ขอขอบพระคุณคุณพ่อคุณแม่และครอบครัวอันเป็นที่รักที่ช่วยเรื่องทุนการศึกษา ซัพพอร์ตทุกอย่างตั้งแต่เด็กจนโตและเป็นกำลังใจที่สำคัญในการพัฒนาตัวเองมาตลอด ทำให้ดิฉันก้าวพ้นความยากลำบากทั้งหมดเพื่อให้สำเร็จและทำให้ครอบครัวภูมิใจ

ขอขอบคุณ ร.ท.นพ.กวิณ วงศ์ธรรมริน อาจารย์ประจำภาควิชาสรีรวิทยา วิทยาลัยแพทยศาสตร์พระมงกุฎเกล้า ที่ช่วยมาเป็น Domain Expert ในด้านความรู้ทางการแพทย์ของผู้ป่วยความดันโลหิตสูงและโรคแทรกซ้อนอื่นๆ และช่วย Review Attribute ที่สำคัญกับการวินิจฉัย

ขอขอบคุณเพื่อนๆทุกคนรวมไปถึงศิลปินที่ชื่นชอบทุกท่านที่มีส่วนช่วยอย่างมากในการให้กำลังใจในวันที่เหนื่อยล้าและเป็นแรงผลักดันอีกทาง ถ้าไม่มีทุกคนผลงานชิ้นนี้คงไม่สำเร็จลุล่วงได้เช่นกัน

กุลยา วิสุทธิเทพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ.....	ง
สารบัญรูป.....	ช
สารบัญตาราง.....	ฉ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการค้นคว้าอิสระ.....	1
1.3 ขอบเขตของการค้นคว้าอิสระ.....	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	2
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	3
2.1 ทฤษฎีและข้อปฏิบัติพื้นฐานของผู้ป่วยโรคอ้วนและโรคความดันโลหิตสูง.....	3
2.2 การตรวจคัดกรองโรคเบาหวานด้วยการวัดค่า FPG.....	6
2.3 การกำหนดระยะของโรคไตเรื้อรังด้วยค่า eGFR.....	7
2.4 13 เขตสุขภาพของประเทศไทย.....	8
2.5 คำจำกัดความ.....	9
2.6 การรวบรวมข้อมูลผู้ป่วย.....	9
2.7 ตัวแบบที่จะใช้ในการศึกษา.....	13
2.7.1 ทฤษฎีบท.....	13
2.7.2 ขั้นตอนการวิเคราะห์การถดถอยโลจิสติก.....	15
2.7.3 การทำ Dummy Variable.....	15
2.7.4 การวัดประสิทธิภาพด้วยตาราง Confusion Matrix.....	16
2.8 ภาษาและเครื่องมือที่ใช้ในการศึกษา.....	18

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.8.1 Python.....	18
2.8.2 RapidMiner	20
บทที่ 3 วิธีการดำเนินงาน.....	21
3.1 ระเบียบวิธีการดำเนินงาน.....	21
3.1.1 กำหนดหัวข้อ ความเป็นมาและความสำคัญของหัวข้อและสมมติฐาน	22
3.1.2 ดำเนินการขอใช้ข้อมูลผู้ป่วยความดันโลหิตสูง	22
3.1.3 นำข้อมูลที่ได้มาทำการสำรวจและนำเข้าวิธีการ Preprocessing.....	22
3.1.4 นำข้อมูลที่ผ่านการ Preprocessing แล้วเข้าตัวแบบ Prediction	22
3.1.5 Result Visualization	23
3.1.6 วิเคราะห์และสรุปผลการดำเนินงาน	23
3.2 ระยะเวลาการดำเนินงาน.....	23
3.3 ขั้นตอนการ Preprocessing	24
3.3.1 Missing Value	24
3.3.2 Generate Attribute.....	24
3.3.3 Select Attribute.....	27
3.4 ขั้นตอนการ Preprocessing ด้วยภาษา Python.....	29
3.5 การนำข้อมูลเข้าตัวแบบ Logistic Regression	32
3.6 การแสดงผล Visualization	32
บทที่ 4 ผลการดำเนินงาน.....	36
4.1 ลักษณะของกลุ่มตัวอย่าง.....	36
4.2 จำนวนประชากรแบ่งแยกตามการเป็นโรคอ้วน.....	37
4.2.1 Gender Group	37
4.2.2 Age Group.....	38
4.2.3 Smoking	38
4.2.4 Cholesterol (bad fats)	39
4.2.5 HDL (good fats).....	39
4.2.6 Occupation Type	40

4.2.7	Chronic Kidney Disease Stages.....	40
4.2.8	Diabetes	41
4.2.9	Area Group	41
4.2.10	Abdominal Obesity.....	42
4.3	อัตราการเปลี่ยนแปลงของ BMI Group ในแต่ละปี	43
4.4	อัตราการเป็นโรคอ้วนในผู้ป่วยความดันโลหิตสูงที่แบ่งโดยเพศในแต่ละปี.....	44
4.5	อัตราการเป็นโรคอ้วนในผู้ป่วยความดันโลหิตสูงที่แบ่งโดยอายุในแต่ละปี	45
4.6	อัตราการเป็นโรคอ้วนของผู้ป่วยความดันโลหิตสูงแบ่งตาม 13 เขตสุขภาพ.....	46
4.7	การหาค่าทางสถิติเบื้องต้น	47
4.8	การหา correlation ระหว่างตัวแปรอิสระ	47
4.9	การแสดงผลข้อมูลหลังจากนำเข้าตัวแบบ Logistic Regression.....	48
4.10	ตาราง Confusion Matrix.....	49
4.11	ค่า Mean Square Error และค่า Root Mean Square Error	49
4.12	Receiver operating characteristic curve (ROC Curve)	50
4.13	การแปลผลตัวแปรอิสระ (Multivariable factors) ที่สัมพันธ์ต่อตัวแปรตาม	51
บทที่ 5	สรุปผลการดำเนินงานและข้อเสนอแนะ	54
5.1	สรุปผลการดำเนินงาน	54
5.2	ข้อจำกัดของการค้นคว้าอิสระ	54
5.3	ข้อเสนอแนะ.....	55
	เอกสารอ้างอิง	56
	ประวัติผู้เขียน.....	59

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 ความชุกของภาวะอ้วน (BMI \geq 25 kg/m ²)	5
2.2 ความชุกของภาวะอ้วนในผู้ใหญ่สังคมนบที่มีแนวโน้มสูงขึ้นเมื่อเปรียบเทียบกับโดยช่วงอายุ.....	5
2.3 ความชุกของภาวะอ้วนในผู้ใหญ่สังคมนบที่มีแนวโน้มสูงขึ้นเมื่อเปรียบเทียบกับโดยเพศ.....	6
2.4 แสดงระยะของโรคไตเรื้อรังเทียบกับความสามารถในการกรองของไต (eGFR).....	7
2.5 ฟังก์ชันโลจิสติก (Logistic Function).....	13
2.6 ตัวอย่างการ import Library.....	19
2.7 ตัวอย่างการ import file และสร้าง Dummy Variable.....	19
2.8 ตัวอย่างการแสดงผล Visualization บน Google Colab	19
2.9 ตัวอย่างใช้โปรแกรม Rapid Miner.....	20
3.1 ระเบียบวิธีการดำเนินงาน.....	21
3.2 Function “Replace Missing Values”.....	24
3.3 Function “Generate Attribute”.....	25
3.4 Function “Select Attribute”.....	27
3.5 Function “remove attribute”.....	27
4.1 Gender Group.....	37
4.2 Age Group.....	38
4.3 Smoking.....	38
4.5 HDL (good fats).....	39

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
4.4 Cholesterol (bad fats).....	39
4.6 Occupation Type.....	40
4.7 Chronic Kidney Disease Stages.....	40
4.8 Diabetes.....	41
4.9 Area Group.....	41
4.10 Abdominal Obesity.....	42
4.11 กราฟแสดงอัตราการเปลี่ยนแปลงของ BMI Group.....	43
4.12 กราฟแสดงอัตราการเป็นโรคอ้วนในผู้ป่วยความดันโลหิตสูงที่แบ่งโดยเพศ.....	44
4.13 กราฟแสดงอัตราการเป็นโรคอ้วนในผู้ป่วยความดันโลหิตสูงที่แบ่งโดยอายุ.....	45
4.14 กราฟแสดงอัตราการเป็นโรคอ้วนของผู้ป่วยความดันโลหิตสูงแบ่งตาม 13 เขตสุขภาพ.....	46
4.15 กราฟ Heatmap แสดงความสัมพันธ์กันของตัวแปร.....	47
4.16 ผลการแสดงค่า Accuracy, Precision, Recall และ f1-score ของตัวแบบ.....	49
4.17 ตาราง Confusion Matrix.....	49
4.18 กราฟ ROC Curve.....	50
4.19 พื้นที่ใต้กราฟ ROC Curve.....	50

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 ข้อมูลเกี่ยวกับคนที่มีอายุมากขึ้นจะมีความเสี่ยงที่จะเป็นโรคอ้วนลดลง.....	4
2.2 แสดงข้อมูลของเขตสุขภาพและจังหวัดที่เป็นสมาชิก.....	8
2.3 แสดงการแปลงข้อมูลจาก Categorical เป็น Binominal (0,1).....	15
2.4 ตัวอย่างตาราง Confusion Matrix.....	16
3.1 การดำเนินงาน.....	23
3.2 Function Expression ใน Function “Generate Attribute”.....	25
3.3 Attributes ที่จะนำไปใช้ต่อในขั้นตอนถัดไป.....	28
4.1 ข้อมูลลักษณะทั่วไปของกลุ่มตัวอย่าง.....	36
4.2 ค่าทางสถิติเบื้องต้นสำหรับข้อมูลที่เป็น Numerical.....	47
4.3 Associated Multivariable Factors.....	53

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

โรคความดันโลหิตสูงเป็นโรคที่ไม่ติดต่อชนิดเรื้อรังที่มีผู้ป่วยจำนวนมากจนสามารถเรียกได้ว่าเป็นปัญหาทางสาธารณสุขระดับโลกในปัจจุบัน โดยในปีพ.ศ. 2558 องค์การอนามัยโลก (World Health Organization: WHO) ระบุว่ามียุโรปผู้ป่วยความดันโลหิตอย่างน้อย 970 ล้านคนทั่วโลกและมีแนวโน้มที่จะสูงขึ้นอย่างต่อเนื่องจนถึง 1,560 ล้านคนในปีพ.ศ. 2568[1]

การเพิ่มของความดันโลหิตที่เป็นสาเหตุของโรคความดันโลหิตสูงเป็นปัจจัยเสี่ยงอิสระที่จะก่อให้เกิดโรคที่ไม่ติดต่ออีกมากมายอาทิเช่นโรคหัวใจขาดเลือด โรคไตเรื้อรัง โรคเส้นเลือดสมองทั้งแบบเลือดออกในสมองและสมองขาดเลือด รวมถึงภาวะหัวใจวาย โดยเฉพาะอย่างยิ่งโรคหัวใจและหลอดเลือดที่มีความสำคัญและอาจนำไปสู่การเสียชีวิตก่อนวัยอันควร แต่ในขณะเดียวกันโรคความดันโลหิตสูงก็ยังเป็นโรคที่สามารถป้องกันได้

โดยปกติการรักษาโรคความดันโลหิตสูงจะทำได้อยู่ 2 วิธีคือการปรับเปลี่ยนวิถีการดำเนินชีวิตหรือการให้ยาลดความดันโลหิต ซึ่งวิธีการปรับเปลี่ยนวิถีการดำเนินชีวิตนั้นทำได้หลากหลายอาทิเช่น การลดน้ำหนักในผู้ที่มีน้ำหนักเกินหรือดัชนีมวลกายมากกว่า 22.9 กก./ตร.กม. จัดการปริมาณเกลือและโซเดียมที่รับประทานต่อวัน ออกกำลังกายอย่างสม่ำเสมอและลดการดื่มแอลกอฮอล์ เป็นต้น

ดังนั้นจึงเป็นที่มาของการค้นคว้าอิสระเพื่อที่จะแสดงแนวโน้มของโรคอ้วนในกลุ่มประชากรชาวไทยที่กำลังได้รับการรักษาโรคความดันโลหิตสูงในโรงพยาบาลเพื่อสะท้อนให้เห็นถึงปัญหาและประสิทธิภาพการส่งเสริมสุขภาพโดยภาพรวมภายในประเทศ อีกทั้งต้องการก่อให้เกิดความตระหนักรู้ในกลุ่มผู้ป่วยและค้นหาปัจจัยที่สัมพันธ์ต่อการควบคุมน้ำหนักเพื่อนำไปสู่การพัฒนานโยบายการควบคุมโรคความดันโลหิตสูงต่อไป

1.2 วัตถุประสงค์ของการค้นคว้าอิสระ

- 1) เพื่อศึกษาแนวโน้มความชุกของโรคอ้วนในผู้ป่วยความดันโลหิตสูงในประเทศไทยปีพ.ศ. 2557 พ.ศ. 2558 และ พ.ศ. 2561
- 2) เพื่อศึกษาปัจจัยร่วมอื่นๆที่มีความสัมพันธ์กับโรคอ้วนและโรคที่ไม่ติดต่อที่เกี่ยวข้องกันในประเทศไทยปีพ.ศ. 2557 พ.ศ. 2558 และ พ.ศ. 2561
- 3) เพื่อหาปัจจัยที่สัมพันธ์กับการควบคุมน้ำหนักในผู้ป่วยความดันโลหิตสูงในประเทศไทยปีพ.ศ. 2557 พ.ศ. 2558 และ พ.ศ. 2561

1.3 ขอบเขตของการค้นคว้าอิสระ

- 1) ในการศึกษาครั้งนี้มุ่งเน้นไปที่การศึกษาหาแนวโน้มความชุกของโรคอ้วนในผู้ป่วยความดันโลหิตสูงในประเทศไทยปีพ.ศ. 2557 พ.ศ. 2558 และ พ.ศ. 2561 และปัจจัยเสี่ยงร่วมอื่นๆ
- 2) ใช้ข้อมูลจากสำนักงานหลักประกันสุขภาพแห่งชาติ (สปสช.) ที่เก็บข้อมูลจากแบบสอบถามของผู้ป่วยโรคความดันโลหิตสูงที่กำลังรักษาตัวในโรงพยาบาลช่วงปีพ.ศ. 2557 พ.ศ. 2558 และ พ.ศ. 2561 แบบไม่ต่อเนื่อง
- 3) ตัวแปรอิสระ (Independent variables) ที่ใช้ศึกษา ได้แก่ น้ำหนัก ส่วนสูง ค่าดัชนีมวลกาย เพศ อายุ อาชีพ ภูมิภาค การนับถือศาสนา สัญชาติ สิทธิการรักษา การสูบบุหรี่หรือการดื่มสุรา ระดับการศึกษา โรคร่วมอื่นๆ ประวัติการใช้ยา (ยารักษาโรคเบาหวาน และยารักษาโรคความดันโลหิตสูง) ผลเลือดที่เคยได้รับการตรวจ
- 4) ตัวแปรตาม (Dependent variables) ที่ใช้ศึกษาได้แก่ การเป็นโรคอ้วน
- 5) ใช้ตัวแบบ Logistic regression ในการคาดการณ์แนวโน้มความชุกและปัจจัยเสี่ยงของโรคอ้วนในผู้ป่วยความดันโลหิตสูงในประเทศไทย

1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ทราบถึงแนวโน้มความชุกของโรคอ้วนในผู้ป่วยความดันโลหิตสูงในประเทศไทยปีพ.ศ. 2557 พ.ศ. 2558 และ พ.ศ. 2561 และปัจจัยเสี่ยงร่วมอื่นๆ
- 2) ทราบข้อมูลการคาดการณ์ที่แสดงออกมาเพื่อจะนำไปสู่การวิเคราะห์สถานการณ์ในอนาคตของโรคดังกล่าวและภาพรวมในประเทศไทย
- 3) สามารถนำข้อมูลเหล่านี้ไปใช้พัฒนาแผนสำหรับการรับมือและรักษาผู้ป่วยโรคความดันโลหิตสูงที่เสี่ยงจะเป็นโรคอ้วนร่วมด้วยในอนาคต

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีและข้อปฏิบัติพื้นฐานของผู้ป่วยโรคอ้วนและโรคความดันโลหิตสูง

โรคอ้วนเกิดขึ้นได้จากหลายปัจจัย สามารถเกิดขึ้นจากการรับประทานอาหารเกินความจำเป็นของร่างกายทำให้ไขมันส่วนที่เกินไปสะสมตามกล้ามเนื้อและชั้นผิวหนังต่างๆ หรือในบางกรณีเกิดจากฮอร์โมนที่ช่วยในการควบคุมการอยากอาหารทำงานผิดปกติ (Insulin) หรือเกิดจากพฤติกรรมมารับประทานที่ผิดปกติเช่นทานไม่หยุดหรือไม่รู้สึกลิ่ม แต่ไม่ว่าสาเหตุของโรคอ้วนจะเกิดขึ้นเพราะเหตุใด โรคอ้วนก็จะเป็นสาเหตุหลักๆที่จะนำผู้ป่วยไปสู่โรคไม่ติดต่อชนิดเรื้อรังอื่นๆเช่นโรคเบาหวาน โรคความดัน และโรคหัวใจ เป็นต้น เนื่องจากระบบต่างๆในร่างกายไม่สามารถทำงานได้อย่างปกติได้และเกิดความไม่สมดุลขึ้นไม่ว่าจะเป็นระดับฮอร์โมน ลักษณะทางกายภาพ น้ำหนักที่มากเกินไปอาจจะส่งผลต่อกระดูกและข้อของผู้ป่วยและสามารถลุกลามไปถึงเกิดผลกระทบต่อสภาวะจิตใจ[2]

8 ใน 10 อันดับแรกของสาเหตุการตายของคนไทยในปีพ.ศ. 2558[3] เกิดจากโรคเรื้อรังเช่นโรคหัวใจขาดเลือด โรคเส้นเลือดในสมอง โรคความดันและโรคเบาหวาน เป็นต้น การรักษาโรคเรื้อรังเหล่านี้ให้ได้ประสิทธิภาพต้องประเมินควบคู่ไปกับโรคอ้วนและภาวะน้ำหนักเกินจึงจะสามารถรักษาได้อย่างถูกต้อง

รายงานเรื่องโรคอ้วนที่ส่งผลต่อความดันโลหิตสูงเป็นที่นิยมมากขึ้นทั่วโลก โดยมีทฤษฎีทางการแพทย์ที่สนับสนุนหลายทฤษฎีอาทิเช่น 1) การเพิ่มขึ้นของฮอร์โมน Insulin ซึ่งจะมีผลต่อไตทำให้ไตดูดกลับเกลือเพิ่มมากขึ้นเป็นสาเหตุที่ทำให้เกิดความดันโลหิตสูงมากขึ้น 2) การสร้าง Leptin ที่มากขึ้นซึ่ง Leptin จะเป็น Polypeptide ที่ทำหน้าที่ในการยับยั้งความอยากอาหารและการผลิตที่มากขึ้นก็จะส่งผลต่อความดันโลหิตสูงได้และ 3) Renin-angiotensin-aldosterone system (RAAS) จะมีการทำงานมากขึ้นในผู้ป่วยโรคอ้วนซึ่งปริมาณ Angiotensin II ที่เพิ่มขึ้นจะส่งผลให้มีระดับความดันโลหิตที่เพิ่มมากขึ้นได้เช่นกัน[4]

การดูแลตัวเองของผู้ป่วยความดันสูงที่เกี่ยวข้องกับการลดค่า BMI เพื่อหลีกเลี่ยงโรคอ้วนประกอบไปด้วย

1) การลดการรับประทานอาหารเค็มโดยบริโภคเกลือไม่เกิน 6 กรัมต่อวัน[4] มีรายงานว่าระดับโซเดียมในปัสสาวะที่สูงขึ้นมีความสัมพันธ์เชิงบวกกับโรคอ้วนอย่างไม่มีนัยสำคัญทางสถิติ เนื่องจากการบริโภคอาหารเค็มทำให้มีโซเดียมในพลาสมาเพิ่มขึ้นซึ่งจะเพิ่มการเคลื่อนย้ายของน้ำจากในเซลล์ไปสู่ช่องว่างระหว่างเซลล์รวมถึงมีการกระตุ้นการกระหายน้ำ และเพิ่มความดันในหลอดเลือดแดง ซึ่งจะมีผลให้ความดันโลหิตและปริมาณเลือดที่ออกจากหัวใจใน 1 นาทีเพิ่มสูงขึ้นและเกิดอาการบวมน้ำทำให้น้ำหนักตัวเพิ่มขึ้น

2) การลดน้ำหนักตัว ต้องควบคุมมวลกายให้อยู่ในระหว่าง 18.5–24.9 kg./m² ควบคุมอาหาร เพิ่มการออกกำลังกายและการทำพฤติกรรมบำบัดโดยความดันโลหิตจะลดลง 5-20 มม.ปรอท เมื่อน้ำหนักตัวลดลง 10 กิโลกรัม

3) การลดความเครียดด้วยวิธีที่เหมาะสม จากรายงานพบว่าผู้ป่วยกลุ่มที่เข้าโปรแกรมการฝึกการจัดการความเครียดมีการพัฒนาความสามารถด้านจิตสังคมที่ดีขึ้นและส่งผลให้การทำงานของหัวใจและหลอดเลือดดีกว่ากลุ่มที่ได้รับการดูแลตามปกติรวมถึงมีความกดดันทางด้านอารมณ์น้อยกว่ากลุ่มที่ได้รับการดูแลตามปกติ[5]

นอกเหนือจากปัญหาการทำงานผิดปกติของสารเคมีในร่างกายที่ส่งผลต่อการเพิ่มขึ้นของน้ำหนักตัวและโรคความดันโลหิตสูงแล้วยังมีปัจจัยภายในที่ทำให้เกิดโรคอ้วนได้ดังนี้ 1) กรรมพันธุ์ ยังไม่ทราบถึงความสัมพันธ์ที่แน่ชัดแต่จากรายงานพบว่าคนที่มิบุคคลในครอบครัวอยู่ในภาวะอ้วนหรือน้ำหนักเกินจะส่งผลให้คนในครอบครัวคนอื่นๆมีค่าดัชนีมวลกายมากกว่าคนปกติมากถึง 40-70% ทั้งในเด็กและผู้ใหญ่ 2) อายุ โดยรวมแล้วอายุที่เพิ่มเติมจะทำให้อัตราการเผาผลาญลดลงและขาดการออกกำลังกาย แต่ในรายงานบางพื้นที่ในชนบทกลับตรงกันข้ามกลุ่มตัวอย่างที่อายุที่เพิ่มขึ้น 1 ปีกลับมีอัตราเสี่ยงที่จะเป็นโรคอ้วนลดลง 1% อาจมีสาเหตุมาจากการทำงานที่ยังต้องใช้แรงงานและการบริโภคโดยจะแสดงผลในตารางที่ 2.1 3) อัตราการเผาผลาญโดยทั่วไปการรับประทานอาหารจะให้พลังงานความร้อนแก่ร่างกายอยู่ที่ประมาณ 10% แต่การมีโรคประจำตัวเช่นโรคไทรอยด์อาจจะทำให้การผลิตความร้อนมีประสิทธิภาพลดลงและเกิดพลังงานหลงเหลือในร่างกายสะสมเป็นไขมันได้มากขึ้น[6]

ตารางที่ 2.1 ข้อมูลเกี่ยวกับคนที่มีอายุมากขึ้นจะมีความเสี่ยงที่จะเป็นโรคอ้วนลดลง

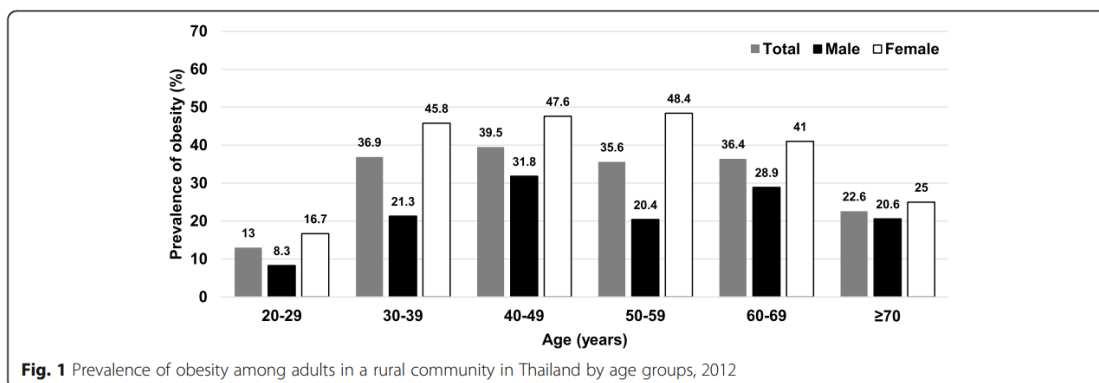
Table 2 Multivariable analysis factors associated with obesity among adults in a rural community in 2018 (n = 627)

Factors	BMI < 25 kg/m ²	BMI ≥ 25 kg/m ²	Adjusted Odds Ratio	95% CI	p-value
Age (years)	56.2 ± 14.9	53.4 ± 11.8	0.99	0.97–0.99	0.029

จากการสำรวจภาวะสุขภาพของประชากรโลกพบว่าภาวะน้ำหนักเกินและโรคอ้วนเป็นปัญหาสุขภาพที่สำคัญและมีแนวโน้มเพิ่มสูงขึ้นอย่างต่อเนื่องตั้งแต่ปีพ.ศ. 2523 เป็นต้นมา และจากการสำรวจครั้งล่าสุดในปี พ.ศ. 2557 ประชากรโลกวัยผู้ใหญ่พบว่าภาวะน้ำหนักเกินมาตรฐานประมาณ 1,900 ล้านคน หรือคิดเป็น 39% และในจำนวนนี้พบว่ามีผู้เป็นโรคอ้วนมากถึง 600 ล้านคน หรือคิดเป็น 13%[6]

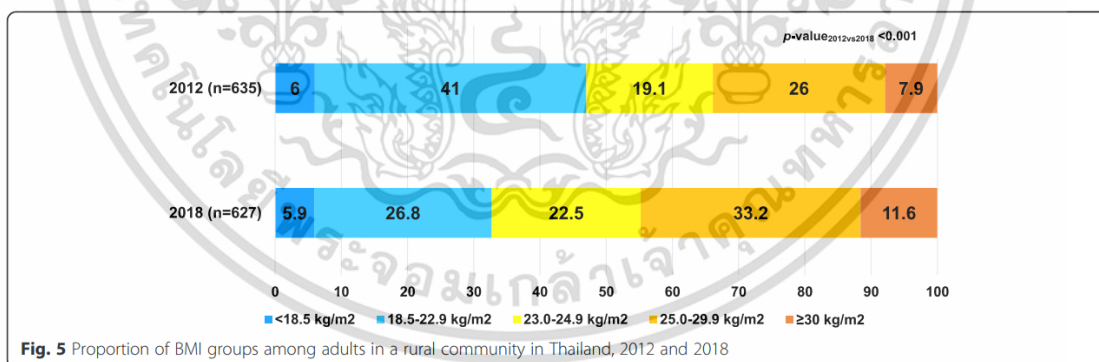
ประเทศไทยในปีพ.ศ. 2557 มีผลรายงานว่าพบคนเป็นโรคอ้วน (BMI ≥ 25 kg/m²) ในประชากรไทยที่อายุ 15 ปีขึ้นไป เป็นเพศชาย 32.9% เป็นเพศหญิง 41.8% เฉลี่ยรวม 37.5%[7] และมีแนวโน้มเพิ่มมากขึ้นในสังคมชนบทและสูงกว่าค่าเฉลี่ยรวมแสดงผลในรูปที่ 2.1 ในรายงานพบอีกว่าโรคอ้วนสัมพันธ์กับค่าความดันโลหิตในระดับสูงและเมื่อสามารถควบคุมและหยุดกระบวนการที่ทำให้เกิดโรคอ้วนได้ภาวะแทรกซ้อนที่เป็นโรคที่ไม่ติดต่อชนิดเรื้อรังต่างๆเหล่านี้ก็จะยุติลงเช่นกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



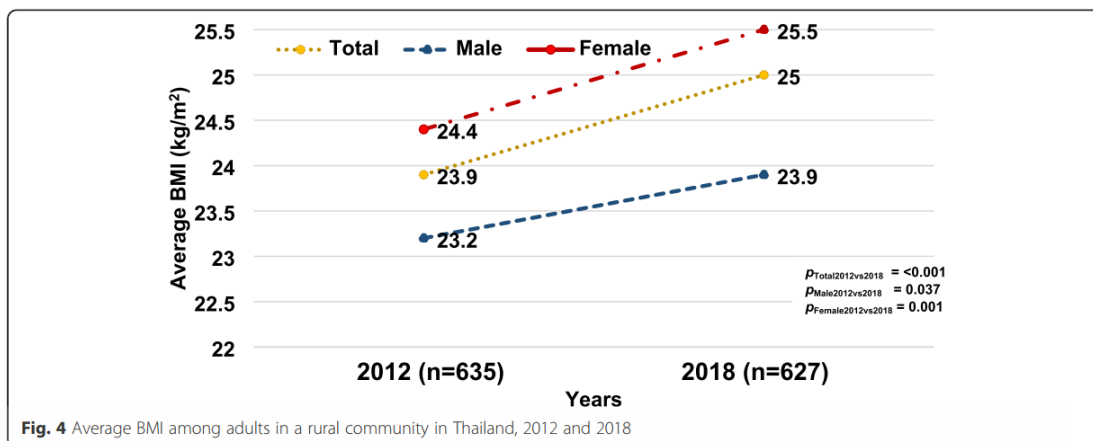
รูปที่ 2.1 ความชุกของภาวะอ้วน (BMI ≥ 25 kg/m²)

ปรากฏการณ์การเพิ่มขึ้นของภาวะอ้วนในผู้ใหญ่อายุระหว่าง 30–90 ปีจากสังคมชนบทและสังคมเมืองอย่างต่อเนื่องดังผลลัพธ์ในรูปที่ 2.2 และ รูปที่ 2.3 นี้สืบเนื่องมาจากการปรับเปลี่ยนไปสู่สังคมอุตสาหกรรมอย่างรวดเร็ว ส่งผลให้เกิดการพัฒนาเศรษฐกิจ ผู้คนเข้าถึงอาหารจากอุตสาหกรรมแปรรูปได้มากขึ้นโดยอาหารเหล่านี้จะเป็นอาหารที่ให้พลังงานสูงเกินจำเป็นและนำไปสู่การเป็นโรคอ้วนได้ [6] ปัจจัยทางอ้อมอีกอย่างหนึ่งคือการที่สังคมเข้ายุคดิจิทัลหรือที่เรียกว่า “การสื่อสารไร้พรมแดน” ทำให้คนขยับร่างกายน้อยลง ในปีพ.ศ. 2559 พบว่าประชาชนใช้ชีวิตติดหน้าจอมือถือหรือหน้าจอทีวีเฉลี่ยวันละ 14 ชั่วโมงและมีการใช้กิจกรรมทางกายน้อยลงซึ่งส่งผลโดยตรงกับการมีภาว่น้ำหนักเกินและโรคอ้วนจึงเป็นหนึ่งในสาเหตุหลักของโรคไม่ติดต่อเรื้อรังอื่นๆ โดยเฉพาะโรคความดันโลหิตสูง



รูปที่ 2.2 ความชุกของภาวะอ้วนในผู้ใหญ่สังคมชนบทมีแนวโน้มสูงขึ้นเมื่อเปรียบเทียบโดยช่วงอายุ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.3 ความชุกของภาวะอ้วนในผู้ใหญ่สังคมชนบทมีแนวโน้มสูงขึ้นเมื่อเปรียบเทียบกับเพศ

อย่างไรก็ตามไม่ว่าจะเป็นสังคมเมืองหรือชนบทก็ยังมีแนวโน้มของโรคอ้วนที่สูงขึ้นอย่างต่อเนื่องและเป็นปัญหาระดับโลกที่ต้องได้รับการแก้ไขอย่างเร่งด่วน การสาธารณสุขจึงควรจัดให้มีการดำเนินการด้านสุขภาพโดยเฉพาะอย่างยิ่งการปรับเปลี่ยนพฤติกรรมกรรมการบริโภคอาหารในชุมชนเพื่อลดค่าดัชนีมวลกาย ปัจจัยเสี่ยงที่นำไปสู่โรคอ้วนควรถูกระงับเพื่อยับยั้งการลุกลามของโรคไม่ติดต่อชนิดต่างๆและภาวะแทรกซ้อนอื่นๆด้วยเพื่อลดอัตราการเสียชีวิตจากโรคเรื้อรังต่างๆที่มีสาเหตุมาจากโรคอ้วนดังกล่าว[7]

2.2 การตรวจคัดกรองโรคเบาหวานด้วยการวัดค่า FPG (Fasting Plasma Glucose)

โดยปกติการตรวจคัดกรองโรคเบาหวานจะใช้วิธีการตรวจวัดจากระดับน้ำตาลในเลือดทั้งหมด 2 ครั้งและต้องมีผลเลขเกณฑ์ทั้ง 2 ครั้ง โดยจะแบ่งเป็น 3 กรณีคือ 1) การเจาะที่ปลายนิ้วขณะอดอาหารด้วยการงดอาหารล่วงหน้าอย่างน้อย 8 ชั่วโมง ถ้าหากค่า FPG มีค่ามากกว่าหรือเท่ากับ 126 มก./ดล. ถือว่าเป็นผู้ป่วยเบาหวาน แต่ถ้าหากน้อยกว่านั้นและอยู่ในช่วง 100 – 125 มก./ดล. ยังไม่ถือว่าเป็นผู้ป่วยเบาหวานและควรตรวจซ้ำอย่างน้อยทุกๆ 1 ปีเนื่องจากมีภาวะน้ำตาลในเลือดผิดปกติ 2) การตรวจสอบความทนต่อน้ำตาลกลูโคส โดยให้ดื่มสารละลายน้ำตาลกลูโคสในระดับที่กำหนด ถ้าหากผ่านไป 2 ชั่วโมงแล้ว ระดับน้ำตาลในเลือดยังสูงเกิน 200 มก./ดล. ก็จะเป็นผู้ป่วยเบาหวาน 3) หากผู้ป่วยมีอาการของโรคเบาหวานแล้ว ให้เจาะเลือดตรวจวัดระดับน้ำตาลในเลือดแบบไม่ต้องอดอาหาร ถ้าหากระดับน้ำตาลในเลือดสูงเกิน 200 มก./ดล. ถือว่าเป็นผู้ป่วยเบาหวานเช่นกัน[8]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 การกำหนดระยะของโรคไตเรื้อรังด้วยค่า eGFR (Estimated Glomerular Filtration Rate)

การแบ่งระยะของโรคไตเรื้อรังแบ่งได้เป็น 2 กรณีคือ กรณีแรกแบ่งระยะของโรคไตตามสาเหตุและตามโรคร่วม เช่น โรคเบาหวาน โรคความดันโลหิตสูง โรคพันธุกรรม โรคที่เกิดจากปัจจัยสิ่งแวดล้อมและโครงสร้างทางกายวิภาคของไต เป็นต้น ส่วนกรณีถัดมาเป็นกรณีที่ทางผู้วิจัยใช้เป็นกรณีสำหรับคัดกรองระยะของผู้ป่วยโรคไตเรื้อรังคือการวัดจากระดับค่า eGFR ซึ่งเป็นค่าที่บ่งบอกถึงระดับความสามารถในการกรองของเสียของไต ในผู้ที่มีความเสี่ยงต่ำจะอยู่ที่ระดับมากกว่า 90% หรือช่วง G1 ระยะถัดมาคือผู้ป่วยโรคไตระยะ G2 จะมีค่า eGFR จะอยู่ในช่วง 60 – 89% ระยะถัดมาคือผู้ป่วยโรคไตระยะ G3a และ G3b จะมีค่า eGFR จะอยู่ในช่วง 30 – 59% จัดได้ว่าเป็นผู้ป่วยมีความเสี่ยงปานกลาง ระยะถัดมาคือผู้ป่วยโรคไตระยะ G4 จะมีค่า eGFR จะอยู่ในช่วง 15 – 29% จัดได้ว่าเป็นผู้ป่วยมีความเสี่ยงสูง และผู้ป่วยโรคไตระยะ G5 จะมีค่า eGFR ต่ำว่า 15% คือผู้ป่วยไตวายระยะสุดท้าย[9] ดังที่แสดงในรูปที่ 2.4

STAGES OF CHRONIC KIDNEY DISEASE	GFR*	% OF KIDNEY FUNCTION
Stage 1 Kidney damage with normal kidney function	90 or higher	90-100%
Stage 2 Kidney damage with mild loss of kidney function	89 to 60	89-60%
Stage 3a Mild to moderate loss of kidney function	59 to 45	59-45%
Stage 3b Moderate to severe loss of kidney function	44 to 30	44-30%
Stage 4 Severe loss of kidney function	29 to 15	29-15%
Stage 5 Kidney failure	Less than 15	Less than 15%

* Your GFR number tells you how much kidney function you have. As kidney disease gets worse, the GFR number goes down.

รูปที่ 2.4 แสดงระยะของโรคไตเรื้อรังเทียบกับความสามารถในการกรองของไต (eGFR)[9]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4 13 เขตสุขภาพของประเทศไทย

ในปัจจุบัน การแบ่งเขตสุขภาพในประเทศไทยของการดูแลประชาชนทางด้านสาธารณสุขจะประกอบไปด้วยทั้งหมด 13 เขต (รวมกรุงเทพมหานคร) มุ่งเน้นพัฒนาระบบบริการทุกระดับขั้นและก่อตั้งศูนย์ความเชี่ยวชาญระดับสูงเพื่อสร้างระบบสาธารณสุขที่เชื่อมโยงเป็นเครือข่ายโดยใช้หลักการ “เขตสุขภาพที่ไร้รอยต่อ (Seamless Health Service Network)” โดยมีการจัดเขตสุขภาพกันดังในตารางที่ 2.2

ตารางที่ 2.2 แสดงข้อมูลของเขตสุขภาพและจังหวัดที่เป็นสมาชิก

ลำดับเขตสุขภาพ	ชื่อเขตสุขภาพ	จังหวัดที่เป็นสมาชิก
เขตสุขภาพที่ 1	เชียงใหม่	เชียงใหม่ เชียงราย แพร่ น่าน พะเยา ลำปาง ลำพูน แม่ฮ่องสอน
เขตสุขภาพที่ 2	พิษณุโลก	ตาก เพชรบูรณ์ พิษณุโลก อุตรดิตถ์ สุโขทัย
เขตสุขภาพที่ 3	นครสวรรค์	กำแพงเพชร พิจิตร นครสวรรค์ ชัยนาท อุทัยธานี
เขตสุขภาพที่ 4	สระบุรี	สระบุรี นนทบุรี ลพบุรี อ่างทอง นครนายก สิงห์บุรี พระนครศรีอยุธยา ปทุมธานี
เขตสุขภาพที่ 5	ราชบุรี	เพชรบุรี สมุทรสาคร สมุทรสงคราม ประจวบคีรีขันธ์ สุพรรณบุรี นครปฐม ราชบุรี กาญจนบุรี
เขตสุขภาพที่ 6	ระยอง	สระแก้ว ปราจีนบุรี ฉะเชิงเทรา สมุทรปราการ ชลบุรี จันทบุรี ระยอง ตราด
เขตสุขภาพที่ 7	ขอนแก่น	กาฬสินธุ์ ขอนแก่น มหาสารคาม ร้อยเอ็ด
เขตสุขภาพที่ 8	อุดรธานี	อุดรธานี สกลนคร นครพนม เลย หนองคาย หนองบัวลำภู บึงกาฬ
เขตสุขภาพที่ 9	นครราชสีมา	ชัยภูมิ นครราชสีมา บุรีรัมย์ สุรินทร์
เขตสุขภาพที่ 10	อุบลราชธานี	อุบลราชธานี ศรีสะเกษ มุกดาหาร อำนาจเจริญ
เขตสุขภาพที่ 11	สุราษฎร์ธานี	นครศรีธรรมราช สุราษฎร์ธานี ภูเก็ต กระบี่ พังงา ระนอง ชุมพร
เขตสุขภาพที่ 12	สงขลา	สงขลา สตูล ตรัง พัทลุง ปัตตานี ยะลา นราธิวาส
เขตสุขภาพที่ 13	กรุงเทพมหานคร	กรุงเทพมหานคร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยในกลุ่มจังหวัดที่อยู่ในเขตเดียวกันจะมีความเชื่อมโยงกันในด้านภูมิศาสตร์ทางการเมืองและเศรษฐกิจ และมีลักษณะทางกายภาพที่คล้ายคลึงกัน ทำให้สามารถนำข้อมูลของผู้ป่วยมาใช้วิธีการทาง Geospatial โดยแบ่งเป็นเขตสุขภาพทั้ง 13 เขตเพื่อแสดงแนวโน้มโรคอ้วนในผู้ป่วยความดันโลหิตสูงได้อย่างค่อนข้างชัดเจน[10]

2.5 คำจำกัดความ

1) ความดันโลหิตสูง (Hypertension) จาก JNC's eight hypertension guideline หมายถึง ระดับความดันโลหิตซิสโตลิก (systolic blood pressure, SBP) ≥ 140 มม.ปรอทและ/หรือความดันโลหิตไดแอสโตลิก (diastolic blood pressure, DBP) ≥ 90 มม.ปรอท และ/หรือ ใช้ยาความดันโลหิตสูง[11]

2) ความดันโลหิตปกติ หมายถึง ระดับความดันโลหิตซิสโตลิก (systolic blood pressure, SBP) น้อยกว่า 140 มม.ปรอท และความดันโลหิตไดแอสโตลิก (diastolic blood pressure, DBP) น้อยกว่า 90 มม.ปรอท และไม่ใช้ยาความดันโลหิตสูง[11]

3) Thailand National Health Security Office's (NHSO) หมายถึง สำนักงานหลักประกันสุขภาพแห่งชาติ

4) โรคอ้วน (obesity) หมายถึง ผู้ที่มี BMI ≥ 25 kg/m²[12]

5) ภาวะน้ำหนักเกิน (overweight) หมายถึง ผู้ที่มี BMI 23.0 – 24.9 kg/m²[12]

2.6 การรวบรวมข้อมูลผู้ป่วย

ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลแบบทุติยภูมิจากสำนักงานหลักประกันสุขภาพแห่งชาติ (สปสช.) ที่เก็บข้อมูลจากแบบสอบถามของผู้ป่วยโรคความดันโลหิตสูงที่กำลังรักษาตัวในโรงพยาบาล ช่วงปีพ.ศ. 2557 – 2561 แบบไม่ต่อเนื่อง โดยเลือกลักษณะของกลุ่มประชากรที่ใช้ในการศึกษาดังนี้

1) ประชากรเป้าหมาย

ผู้ป่วยความดันโลหิตสูงที่มีอายุมากกว่าหรือเท่ากับ 18 ปีที่ได้รับการตรวจรักษาในโรงพยาบาลสังกัดกระทรวงสาธารณสุขที่อยู่ภายใต้ Thailand National Health Security Office's (NHSO) ยกเว้น โรงพยาบาลของมหาวิทยาลัยหรือวิทยาลัยคณะแพทยศาสตร์

2) การเลือกตัวอย่าง

ทำการเลือกกลุ่มตัวอย่างเพื่อแสดงถึงประชากรความดันโลหิตสูงที่รับการรักษาในโรงพยาบาลทั้งประเทศด้วยวิธีการ Stratified two-stage cluster sampling method with proportional to the size (PPS)

วิธีการ Stratified two-stage cluster sampling method with proportional to the size (PPS) เป็นวิธีการสุ่มกลุ่มตัวอย่างจาก 2 ขั้นตอน ซึ่งในขั้นตอนแรกจะสุ่มแบ่งเป็นกลุ่มใหญ่ๆให้ได้ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จำนวนหนึ่ง นำตัวอย่างที่สุ่มได้ในครั้งแรกมาสุ่มหากลุ่มย่อยๆ และนำตัวอย่างที่สุ่มได้ในครั้งที่ 2 มาศึกษาข้อมูลแบบเจาะลึกยกตัวอย่างเช่น เมื่อสุ่มตลาดได้จำนวนหนึ่งแล้วทำการสุ่มร้านค้าตามจำนวนที่ต้องการจากร้านค้าในตลาดที่สุ่มได้ในครั้งแรกเมื่อสุ่มได้ร้านค้าใดจึง สอบถามราคาสินค้าจากร้านค้าที่สุ่มได้ในขั้นที่สอง โดยจำนวนสุ่มตัวอย่างที่ถูกเลือกมาจะถูกกำหนดความน่าจะเป็นให้เป็นสัดส่วนกับขนาดของประชากร[13],[14]

เกณฑ์การคัดเข้า :

ผู้ป่วยความดันโลหิตสูงที่มีอายุมากกว่าหรือเท่ากับ 18 ปีที่ได้รับการตรวจรักษาในโรงพยาบาลสังกัดกระทรวงสาธารณสุขที่อยู่ภายใต้ Thailand National Health Security Office's (NHSO) program เป็นเวลาเท่ากับหรือนานกว่า 12 เดือนขึ้นไป

เกณฑ์การคัดแยกอาสาสมัคร :

- 1) ผู้ป่วยที่เคยเข้ารับการรักษาทางคลินิก
- 2) ผู้ที่มารับการตรวจในรพ.ของมหาวิทยาลัยหรือวิทยาลัยคณะแพทยศาสตร์
- 3) มีประวัติการตั้งครรภ์ในช่วง 12 เดือนที่ผ่านมา
- 4) ผู้ที่ขาดข้อมูลใน Case record form ใดๆอย่างหนึ่งดังต่อไปนี้

รหัสประจำตัว (PID) วันที่ทำการเก็บข้อมูล เพศ อายุ น้ำหนักล่าสุดภายใน 12 เดือนและส่วนสูง โดยจะใช้ตัวอย่างแบบฟอร์มในการเก็บข้อมูลเกณฑ์การคัดเข้าและเกณฑ์การคัดแยกอาสาสมัครตามรายละเอียดด้านล่าง ดังนี้

ตัวอย่างแบบฟอร์ม

โครงการประเมินผลการดูแลผู้ป่วยโรคเบาหวานชนิดที่2และความดันโลหิตสูงของโรงพยาบาลใน
สังกัดกระทรวงสาธารณสุข และโรงพยาบาลสังกัดกรุงเทพมหานคร ปี 2556

Inclusion Criteria

1. ผู้ป่วยด้วยโรคความดันชนิดที่2และหรือโรคความดันโลหิตสูงที่เข้ามารับบริการในโรงพยาบาล
ด้วยโรคดังกล่าว เป็นเวลามากกว่าหรือเท่ากับ 12 เดือนขึ้นไป yes no
2. อายุตั้งแต่35ปีขึ้นไปที่มีสติสัมปชัญญะสมบูรณ์และไม่มีอาการเจ็บป่วยทางจิต yes no
3. ให้ความยินยอมเข้าร่วมโครงการโดยการลงนามในเอกสารขอความยินยอมจากอาสาสมัครก่อน
เริ่มการวิจัย วันที่เซ็นยินยอม (วัน/เดือน/ปี) ___/___/255___ yes no

Exclusion Criteria

1. ผู้ป่วยโรคเบาหวานชนิดที่1และรักษาด้วยยา Insulin เพียงอย่างเดียว yes no
2. ผู้ป่วยที่เป็นอาสาสมัครที่กำลังอยู่ในโครงการวิจัยเชิงทดลอง yes no
3. ผู้ป่วยที่ขาดนัดติดตามการรักษาในวันที่เก็บข้อมูลและไม่สามารถติดตามตัวให้มารับการรักษาต่อเนื่องได้
 yes no

-
1. ไม่สามารถเข้าร่วมโครงการได้ (ยุติการกรอกข้อมูล)
2. สามารถเข้าร่วมโครงการได้โดยมี รหัสประจำตัวคือ _____
ผู้กรอกข้อมูล _____ Date ___/___/255___
-

การมาติดตามการรักษาครั้งนี้ ตรงตามนัดหมายหรือไม่

1. ตรงตามนัดหมาย 2. ไม่ตรงนัดหมาย

แหล่งที่มาของผู้ป่วยมาจาก OPD ไต (ตอบเพียงข้อเดียวเท่านั้น)

1. OPD เฉพาะโรคเบาหวาน
2. OPD เฉพาะโรคความดันโลหิตสูง
3. OPD เฉพาะโรค**ร่วม**เบาหวาน**และ**ความดันโลหิตสูง
4. OPD อายุรกรรมทั่วไป
5. OPD ตรวจโรคทั่วไป หรือ OPD ประกันสังคม หรือ OPD เวชศาสตร์ครอบครัว
6. OPD อื่นๆ ระบุ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนที่ 1 ข้อมูลทั่วไป

1. เพศ ชาย หญิง
2. อายุ ____ ปี (เช่น พ.ศ. 2556 - พ.ศ.เกิด) ช่วงอายุตั้งแต่ 35-99 ปี
3. อาชีพ (ปัจจุบัน)
 - 1. เกษตรกร, ทำไร่, ทำสวน, ประมง
 - 2. แม่บ้าน, ไม่ได้ประกอบอาชีพ
 - 3. รับจ้างทั่วไป
 - 4. ข้าราชการ
 - 5. นักเรียน/นักศึกษา
 - 6. พนักงาน/ลูกจ้างรัฐวิสาหกิจ
 - 7. ธุรกิจส่วนตัว
 - 8. พระภิกษุ, นักบวช, แม่ชี
 - 9. อื่นๆ ระบุ
 - 10. ไม่มีข้อมูลในเวาระเบียน
4. การนับถือศาสนา
 - 1. พุทธ 2. อิสลาม 3. คริสต์ 4. อื่นๆ ระบุ
5. สิทธิการรักษา
 - 1. สิทธิ 30 บาท ประกันสุขภาพถ้วนหน้า ทหารผ่านศึก
 - 2. เบิกตรง สวัสดิการราชการ ข้าราชการบำนาญ
 - 3. ประกันสังคม
 - 4. สวัสดิการของบริษัทเอกชน องค์กร รัฐวิสาหกิจ
 - 5. จ่ายเงินเอง / ประกันสุขภาพบริษัทเอกชน
 - 6. อื่นๆ ระบุ
6. ประเภทของผู้ป่วย (ตอบเพียงข้อเดียวเท่านั้น)
 - 1. โรคเบาหวานชนิดที่2อย่างเดียว (บันทึกข้อมูลตั้งแต่หน้า 1 ถึง 8)
 - 2. โรคความดันโลหิตสูงอย่างเดียว (บันทึกข้อมูลตั้งแต่หน้า 9 ถึง 11)
 - 3. ได้รับการวินิจฉัยรวมทั้งโรคเบาหวานและความดันโลหิตสูง (บันทึกข้อมูลตั้งแต่หน้า 1 ถึง 11)
7. น้ำหนักล่าสุด _____ กิโลกรัม ไม่มีข้อมูลในเวาระเบียน / ไม่มีการชั่งน้ำหนัก
8. ส่วนสูงล่าสุด _____ เซนติเมตร ไม่มีข้อมูลในเวาระเบียน / ไม่มีการวัดส่วนสูง
9. รอบเอวล่าสุด _____ เซนติเมตร ไม่มีข้อมูลในเวาระเบียน / ไม่มีการวัดรอบเอว

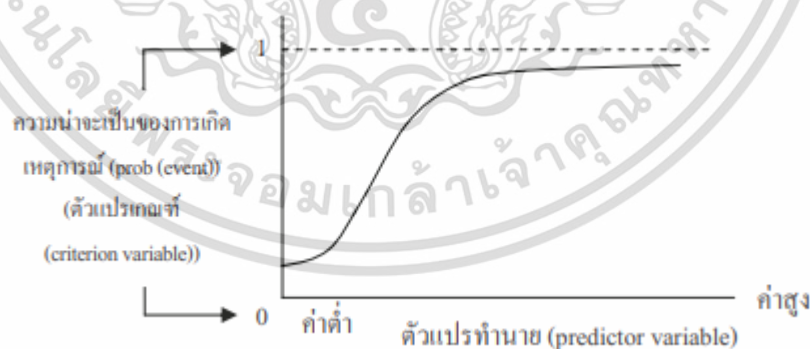
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7 ตัวแบบที่จะใช้ในการศึกษา

2.7.1 ทฤษฎีบท

การศึกษาในครั้งนี้จะใช้ตัวแบบวิเคราะห์การถดถอยโลจิสติก (Logistic regression analysis) เนื่องจากนำเอาข้อมูลดิบไปเข้ากับตัวแบบการ Classification แบบอื่นๆ อาทิเช่น Decision Tree, Random forest และ Artificial Neural Network ทั้ง 3 วิธีจะให้ผล Accuracy ได้น้อยกว่าตัวแบบ Logistic Regression อีกทั้งในงานวิจัยทางการแพทย์อื่นๆ ที่เป็นการพยากรณ์ตัวแปรเชิงคุณภาพ (Qualitative statistical techniques) ก็จะมีนิยมใช้ตัวแบบ Logistic Regression มากกว่าจึงเลือกใช้ตัวแบบนี้มาพยากรณ์ โดยจะแสดงผลกราฟของฟังก์ชันโลจิสติกในรูปที่ 2.5 และจัดอยู่ในหมวดหมู่ของการเรียนรู้แบบมีผู้สอน (Supervised learning)[15]

การวิเคราะห์โลจิสติกถูกใช้อย่างแพร่หลายในวงการแพทย์ เช่นการหาแนวโน้มในการเกิดโรคหนึ่งจากกลุ่มอาการ ปัจจัยเสี่ยงหรือโรคแทรกซ้อนอื่นๆ เนื่องจากมีเป้าหมายคือเพื่อทำนายโอกาสที่จะเกิดเหตุการณ์ที่สนใจ ซึ่งก็คือตัวแปรตามโดยอาศัยสมการโลจิสติกที่สร้างขึ้นจากชุดตัวแปรอิสระ (x's) ที่มีข้อมูลเป็นตัวแปรที่มีข้อมูลอยู่ในระดับช่วง (Interval scale) เป็นอย่างน้อย หากเป็นข้อมูลเชิงกลุ่มจะต้องแปลงเป็นตัวแปรทวิที่มีค่า 0 กับ 1 ก่อน โดยที่ระหว่างตัวแปรอิสระจะต้องมีความสัมพันธ์กันต่ำโดยใช้เกณฑ์ค่า r ไม่เกิน .65 ถ้าใช้เกณฑ์ของ Burns and Grove (1993) หรือถ้าใช้เกณฑ์ของ Stevens (1996) ค่า r ไม่เกิน .80 ซึ่งถ้าหากเกิดความสัมพันธ์กันสูงจะทำให้เกิดปัญหา Multicollinearity และในการวิเคราะห์จะต้องใช้ขนาดตัวอย่างหรือ n มากกว่าหรือเท่ากับ 30 เท่าของจำนวนตัวแปรอิสระโดยจะแบ่ง Dataset ออกเป็น 2 กลุ่มคือ Train set เพื่อใช้ในการ Train ตัวแบบและ Test set ใช้ในการวัดประสิทธิภาพของตัวแบบด้วยอัตรา 7:3 ส่วน[16]



รูปที่ 2.5 ฟังก์ชันโลจิสติก (Logistic Function)

กรณีตัวแปรอิสระมากกว่า 1 ตัว (ตัวแปร $x > 1$ ตัว) ในการวิเคราะห์เมื่อตัวแปรทำนายมีมากกว่า 1 ตัวจะได้ฟังก์ชันดังสมการที่ (2.1)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$P_y = \frac{e^{b_0 + b_1 x_1 + \dots + b_p x_p}}{1 + e^{b_0 + b_1 x_1 + \dots + b_p x_p}} \quad \text{สมการที่ (2.1)}$$

โดยที่

เมื่อ P_y = ความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ

จะได้ Q_y หรือความน่าจะเป็นของการไม่เกิดเหตุการณ์ที่สนใจดังสมการที่ (2.2)

$$Q_y = 1 - P_y$$

หรือ

$$Q_y = 1 - \left(\frac{e^{b_0 + b_1 x_1 + \dots + b_p x_p}}{1 + e^{b_0 + b_1 x_1 + \dots + b_p x_p}} \right) \quad \text{สมการที่ (2.2)}$$

จากความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามของการวิเคราะห์ถดถอยโลจิสติกไม่เป็นรูปเชิงเส้นจึงต้องมีการปรับให้ความสัมพันธ์ให้อยู่ในรูปเชิงเส้นในรูปแบบของ odds หรือ odd ratio

odds หรือ odd ratio หมายถึงอัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์ที่สนใจ ($y = 1$) กับโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ ($y = 0$) หรือจะได้ดังสมการที่ (2.3)

$$\text{odds} = \frac{P_y \text{ (โอกาสเกิดเหตุการณ์ที่สนใจ)}}{Q_y \text{ (โอกาสไม่เกิดเหตุการณ์ที่สนใจ)}} \quad \text{สมการที่ (2.3)}$$

ค่าของ odds จะแสดงถึงโอกาสที่จะเกิดเหตุการณ์ที่สนใจเป็นกี่เท่าของโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจเช่น odds ของการโยนเหรียญ 1 ครั้งเท่ากับ $0.5/0.5 = 1$ หรือถ้า odds มีค่าเท่ากับ 2.5 แสดงว่าโอกาสที่จะเกิดเหตุการณ์ที่สนใจเป็น 2.5 เท่าของโอกาสที่จะไม่เกิดถ้า odds มีค่าเท่ากับ 1 โอกาสที่จะเกิดเหตุการณ์ที่สนใจกับโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจเท่ากันนั่นคือถ้า odds มีค่ามากกว่า 1 แสดงว่าโอกาสที่จะเกิดเหตุการณ์ที่สนใจนั้นมากกว่าโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7.2 ขั้นตอนการวิเคราะห์การถดถอยโลจิสติก มีดังนี้

ขั้นตอนที่ 1 กำหนดตัวแบบสมการการถดถอยถ้าตัวแปรตามมีค่าเพียง 2 ค่าจะกำหนดเป็นแบบ Binary Logistic Regression ถ้ามีตัวแปรตามมากกว่า 2 ค่าจะกำหนดเป็นแบบ Multinomial Logistic Regression

ขั้นตอนที่ 2 ตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์

ขั้นตอนที่ 3 คัดเลือกตัวแปรอิสระเข้าในตัวแบบประกอบด้วยวิธี enter method วิธี Forward stepwise: condition วิธี Backward stepwise: condition วิธี Backward stepwise: likelihood วิธี Forward stepwise: Wald วิธี Backward stepwise: Wald และวิธี Forward stepwise: likelihood

ขั้นตอนที่ 4 ประมาณค่าสัมประสิทธิ์การถดถอย β_i ด้วยเทคนิคของความน่าจะเป็นสูงสุด (Maximum Likelihood) คำนวณหาค่า Odd Ratio และสร้างสมการ

ขั้นตอนที่ 5 ตรวจสอบความเหมาะสมของตัวแบบวิธีการตรวจสอบความเหมาะสมของตัวแบบที่นิยมใช้มีหลายวิธีซึ่งจะเลือกใช้วิธีใดก็ได้ตามความเหมาะสมของข้อมูล

ขั้นตอนที่ 6 วัดระดับความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระโดยใช้พิจารณาหรือตรวจสอบความสอดคล้องของตัวแบบหรือร้อยละที่สามารถอธิบายความแปรปรวนหรือความผันแปรตัวสถิติที่ใช้

ขั้นตอนที่ 7 นำตัวแบบไปใช้พยากรณ์ตัวแปรตาม[17]

2.7.3 การทำ Dummy Variable

เนื่องจากการที่จะนำข้อมูลเข้า Logistic Regression Model นั้นต้องใช้ข้อมูลเชิงปริมาณเท่านั้นซึ่ง Attributes หลายๆตัวเป็นตัวแปรเชิงคุณภาพ อาทิเช่น เพศ การเป็นหรือไม่เป็นโรคแทรกซ้อนอื่นๆ ทำให้ต้องนำข้อมูล Categorical ดังกล่าวนำมาแปลงเป็น Numerical (0,1) เพื่อที่จะทำให้สามารถนำเข้าตัวแบบได้โดยที่ค่าไม่ผิดเพี้ยนไปจากความหมายเดิม

โดยการใช้ Dummy Variable เข้ามาช่วยซึ่งจะเปลี่ยนชุดข้อมูลแบบ Categorical ให้เป็น Binominal (0,1) และแบ่งตัวแปรออกเป็นชุดข้อมูลดังตัวอย่างตามตารางที่ 2.3[18]

ตารางที่ 2.3 แสดงการแปลงข้อมูลจาก Categorical เป็น Binominal (0,1)

No	gender	No	gender_male	gender_female
1	male	1	1	0
2	female	2	0	1
3	female	3	0	1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7.4 การวัดประสิทธิภาพด้วยตาราง Confusion Matrix

การที่จะสามารถนำตัวแบบไปใช้ได้นั้นต้องมีการนำตัวแบบดังกล่าวมาวัดประสิทธิภาพก่อนว่าตัวแบบมีประสิทธิภาพมากพอที่จะนำมาพัฒนาต่อหรือใช้ได้จริงในงานด้านต่างๆ ซึ่งการวัดประสิทธิภาพที่ได้รับความนิยมและจะถูกกล่าวถึงในการค้นคว้าอิสระนี้คือวิธีการวัดประสิทธิภาพจากการทำ Confusion Matrix ซึ่งเป็นตารางที่วัดความสามารถของค่าจริงและผลจากการทำนายโดยจะประกอบไปด้วยทั้งหมด 4 ช่องดังตารางที่ 2.4[19]

ตารางที่ 2.4 ตัวอย่างตาราง Confusion Matrix

		Actual values	
		Positive (1)	Negative (0)
Predictive values	Positive (1)	TP	FP
	Negative (0)	FN	TN

โดยที่ True Positive(TP) คือสิ่งที่ตัวแบบทำนายว่า “จริง” และมีค่าเป็น “จริง” ตรงตามทำนาย True Negative(TN) คือสิ่งที่ตัวแบบทำนายว่า “ไม่จริง” และมีค่า “ไม่จริง” ตรงตามทำนาย False Positive(FP) คือสิ่งที่ตัวแบบทำนายว่า “จริง” แต่มีค่าเป็น “ไม่จริง” ไม่ตรงตามทำนาย False Negative(FN) คือสิ่งที่ตัวแบบทำนายว่า “ไม่จริง” แต่มีค่าเป็น “จริง” ไม่ตรงตามทำนาย

จากนั้นนำตัวเลขในตารางมาเข้าสมการเพื่อหาค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F1-score) โดยที่

1) Accuracy

สมการแสดงค่าความถูกต้องของตัวแบบโดยวัดจากผลรวมของตัวเลขบนเส้นทแยงมุมในตาราง Confusion Matrix หารด้วยจำนวน observations ทั้งหมดดังสมการที่ (2.4)

$$\frac{TP + TN}{TP + TN + FP + FN}$$

สมการที่ (2.4)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) Precision

สมการแสดงความแม่นยำของตัวแบบเป็นการเปรียบเทียบการทำนายว่าจริง โดยมี 2 กรณี คือ ทำนายว่าจริงแล้วเกิดขึ้นจริง (TP) และทำนายว่าจริงแล้วไม่เกิดขึ้นจริง (FP) ดังสมการที่ (2.5)

$$\frac{TP}{TP + FP} \quad \text{สมการที่ (2.5)}$$

3) Recall

สมการแสดงความถูกต้องของการทำนายว่าจะป็นจริงเทียบกับจำนวนครั้งของเหตุการณ์การทำการทำนายและเกิดขึ้นว่าเป็นจริง ดังสมการที่ (2.6)

$$\frac{TP}{TP + FN} \quad \text{สมการที่ (2.6)}$$

4) F1-score (F-measure)

สมการแสดงความเฉลี่ยแบบ harmonic mean ระหว่าง Precision และ Recall จุดประสงค์ของการสร้าง F1 ขึ้นมาคือเพื่อเป็น single metric ที่วัดความสามารถของตัวแบบดังสมการที่ (2.7)

$$2 \left(\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \right) \quad \text{สมการที่ (2.7)}$$

2.7.5 การหาค่า Mean Square Error และ Root Mean Square Error

1) Mean Square Error (MSE) หรืออีกชื่อหนึ่งเรียกว่า L2 Loss คือการหาค่าพิกัด Error จาก y จริง และ y ทำนาย จะทำให้ค่า error ของทุกจุดเป็นบวกก่อนด้วยการนำไปยกกำลังสอง แล้วนำค่า error ไปหาค่าเฉลี่ย ดังสมการที่ (2.8)[20]

$$MSE = \frac{1}{n} \sum_{i=1}^n (\text{error})^2 \quad \text{สมการที่ (2.8)}$$

2) Root Mean Square Error (RMSE) เป็น Loss Function หนึ่งที่นำเอา MSE มาใส่ Square Root เนื่องจากหน่วยของค่า error จาก MSE เป็นค่ากำลังสอง จึงใส่ Square root เข้าไปเพื่อนำค่าเดิมกลับมา ทำให้ตีความง่ายขึ้น ดังสมการที่ (2.9)[20]

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{error})^2} \quad \text{สมการที่ (2.9)}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับทำซ้ำเพื่อการศึกษานั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.8 ภาษาและเครื่องมือที่ใช้ในการศึกษา

2.8.1 Python

ภาษา Python [21] ถูกใช้ในการศึกษาการคาดการณ์แนวโน้มความชุกและปัจจัยเสี่ยงของโรค อ้วนในผู้ป่วยความดันโลหิตสูงในประเทศไทย โดยภาษา Python เป็นภาษาระดับสูงชนิดหนึ่งที่ถูก ออกแบบให้ Code สามารถอ่านได้ง่ายและใกล้เคียงกับภาษาของมนุษย์โดยใช้ช่องไฟหรือ spacebar ในการบ่งบอกว่า Code บรรทัดนั้นอยู่ภายใต้ Code Block ไตแทนที่จะใช้สัญลักษณ์ปีกกาอย่าง ภาษาอื่นเช่น Java, C++

ภาษา Python มี Library ที่จำเป็นต่อการวิเคราะห์ข้อมูลมากมายเช่น Numpy ที่ใช้ในการ ช่วยคำนวณทางคณิตศาสตร์ Pandas ใช้ในการช่วยจัดการกับ Data frame และในขั้นตอนการ Preparing Data SciKit-Learn คือ Library ที่ช่วยในการทำ Machine Learning เป็นต้น อีกทั้งยังสามารถเรียกใช้ Library สำเร็จรูปจาก Git-hub ที่มีผู้พัฒนาท่านอื่นๆสร้างไว้แล้วมาใช้ร่วมได้และ เนื่องด้วยภาษา Python เป็นภาษายอดนิยมที่ใช้ในการทำงานสาย Data Science จึงมีตัวอย่าง Code และ Document ที่เกี่ยวข้องมากมายที่ทำให้ง่ายต่อการพัฒนาต่อและแก้ไข Error ที่อาจจะ เกิดขึ้นได้ระหว่างการพัฒนา [21]

ภาษา Python กำเนิดขึ้นตั้งแต่ปีค.ศ. 1980 และเริ่มใช้งานกันช่วงธันวาคมปีค.ศ. 1989 โดย นาย Guido van Rossum ภายใต้ Centrum Wiskunde & Informatica (CWI) ที่ Netherlands และหลังจากนั้นหลายปีภาษา Python version 2.0 ถูกปล่อยออกมาให้ใช้ตอน 16 ตุลาคมปีค.ศ. 2000 ซึ่งมี Feature ใหม่ ๆ มากมายเช่น Cycle-detecting, Garbage Collector และรองรับ Unicode ใน version มีผลกระทบต่อ Community จำนวนมากเพราะมีการดำเนินการหลาย อย่างที่เปลี่ยนไป

Python 3.0 ถูกปล่อยออกมาให้ใช้งานตั้งแต่วันที่ 3 ธันวาคมปีค.ศ. 2008 แต่พบปัญหาหลาย อย่างและหลังจากทดสอบอยู่นาน Feature ใหม่ ๆ หลายตัวจึงกลับไปพัฒนาบน Python 2.6 และ Python 2.7 แทนและในปัจจุบัน Python 2.7 คือ version สุดท้ายใน series 2.0 และได้มีการ ประกาศสิ้นสุดการพัฒนา (End of Life) ตั้งแต่ปีค.ศ. 2015 แต่เลื่อนมาเป็นปีค.ศ. 2020 อัน เนื่องมาจาก Code จำนวนมากไม่สามารถ Upgrade ขึ้นไปยัง Python 3.0 ได้ ในช่วงต้นปีค.ศ. 2017 ทาง Google จึงประกาศที่จะช่วยทำการแปลง Library Python 2.7 ให้เพื่อเป็นการสิ้นสุด Python 2.x [22]

ปัจจุบันการพัฒนาภาษา Python สามารถทำได้ทั้งการติดตั้งลงบนเครื่องและการพัฒนาบน Cloud ซึ่งสามารถใช้ได้ฟรีและมีวิธีการติดตั้งที่ง่ายโดยในการค้นหาว่าอิสระนี้จะขอยุติถึงการ พัฒนาบน Cloud เป็นหลักซึ่งมีรายละเอียดดังนี้

การพัฒนาภาษา Python บน Cloud ที่เรียกว่า Colaboratory หรือที่เรียกสั้นๆ ว่า "Colab" ช่วยให้เราพัฒนาและเรียกใช้ Python ผ่านหน้าเว็บไซต์โดยไม่ต้องกำหนดค่าใดๆ สามารถเข้าถึง GPU ได้

โดยไม่มีค่าใช้จ่ายและสามารถแชร์ร่วมกันกับผู้พัฒนาท่านอื่นได้โดยสามารถเข้าไปสร้าง Workspace ได้ที่ลิงค์ <https://colab.research.google.com/> แสดงดังรูปที่ 2.6 – 2.8

Step1: Import Library

```
[ ] #import Library
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn import preprocessing
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from matplotlib import pyplot as plt
import seaborn as sns
import statsmodels.api as sm
from sklearn.decomposition import PCA
from sklearn.metrics import mean_squared_error, roc_curve, roc_auc_score
import math
from sklearn.model_selection import train_test_split
```

/usr/local/lib/python3.7/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning: pandas.util.testing is deprecated. Use the functions in the public API at pandas.testing instead.
import pandas.util.testing as tm

รูปที่ 2.6 ตัวอย่างการ import Library

Step2: Import file csv

```
[ ] #read file
df2 = pd.read_csv("/content/factor_vars.csv")
```

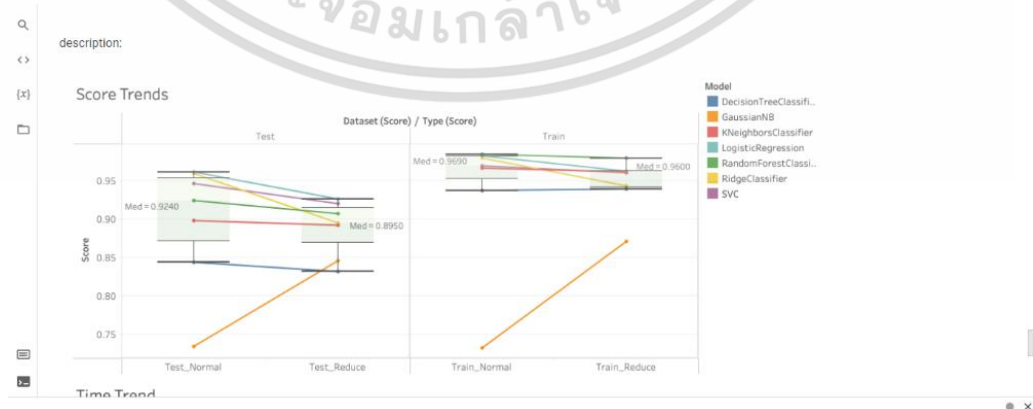
Step3: Preprocessing (dummy var)

```
[ ] #create dummy var
training=pd.get_dummies(df2, columns=["gender", "smoking", "cholesterol", "triglyceride", "hdl", "ldl", "occ_type", "ckd_stage", "diabetes", "age_group", "area_group", "metabolic_synd"])
final_data = training
final_data.head()
```

	age	weight	waist	area	dbp_avg	sbp_avg	obesity	gender_female	gender_male	smoking_no	...	age_group_Elder	age_group_Senior	age_group_Young Adult	area_group_Bangkok	area_group
0	50.0	50.0	16.3	2.0	83.7	127.7	0.0	0	1	1	...	0	1	0	0	0
1	57.0	60.0	19.4	9.0	70.0	135.0	0.0	0	1	1	...	0	1	0	0	0
2	60.0	52.0	19.8	12.0	87.3	142.0	0.0	0	1	0	...	0	1	0	0	0
3	55.0	65.0	20.0	9.0	85.5	135.5	0.0	0	1	0	...	0	1	0	0	0
4	85.0	50.4	20.8	9.0	76.0	156.0	0.0	0	1	0	...	1	0	0	0	0

5 rows x 40 columns

รูปที่ 2.7 ตัวอย่างการ import file และสร้าง Dummy Variable



รูปที่ 2.8 ตัวอย่างการแสดงผล Visualization บน Google Colab

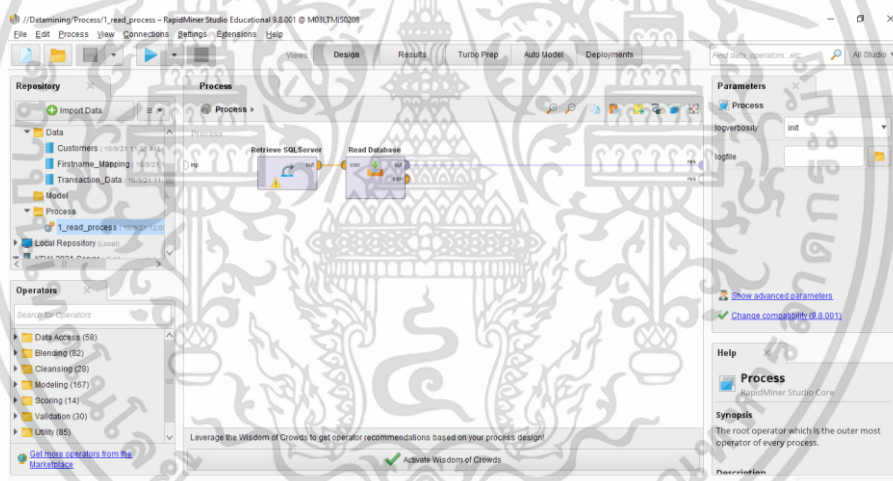
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.8.2 RapidMiner

โปรแกรม RapidMiner[23] คือซอฟต์แวร์ที่ใช้สำหรับการวิเคราะห์ข้อมูลในรูปแบบต่างๆ อาทิเช่นการเตรียมข้อมูล การทำ Machine Learning การทำ Deep Learning การทำ Data Mining และการวิเคราะห์การทำนาย (Predictive analysis) เป็นซอฟต์แวร์ที่ช่วยในการทำ Pipeline ลำดับขั้นการวิเคราะห์ข้อมูลเพื่อลดความเสี่ยงที่จะเกิดข้อผิดพลาดจากการเขียน Code

โปรแกรม RapidMiner สามารถทำการวิเคราะห์ข้อมูลได้อย่างง่ายดายและมีประสิทธิภาพสูง เนื่องจากการทำงานแบบ Drag and Drop โดยที่รูปแบบการนำเสนอในโปรแกรม RapidMiner สามารถนำเสนอ Visualization จากข้อมูลและการวิเคราะห์เชิงพยากรณ์ได้อย่างหลากหลายรูปแบบ รวมถึงการสร้างตัวแบบทางสถิติ ซึ่งในส่วนของการกำหนดค่า Parameter ต่างๆนักวิเคราะห์ข้อมูลสามารถปรับปรุ่ค่าต่างๆใน Parameter ให้ตรงตามรูปแบบที่ต้องการใช้ได้

สำหรับการวิเคราะห์ในขั้นพื้นฐานผู้ใช้งานได้ออกแบบ Workflow ในการวิเคราะห์ข้อมูลในรูปแบบของรูปภาพ ซึ่งทุกอย่างจะเกิดขึ้นในหน้า Design View เพื่อให้ผู้อ่านเข้าใจการใช้งานซอฟต์แวร์ตัวนี้ง่ายขึ้น[23] แสดงผลในรูปที่ 2.9



รูปที่ 2.9 ตัวอย่างใช้โปรแกรม Rapid Miner

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

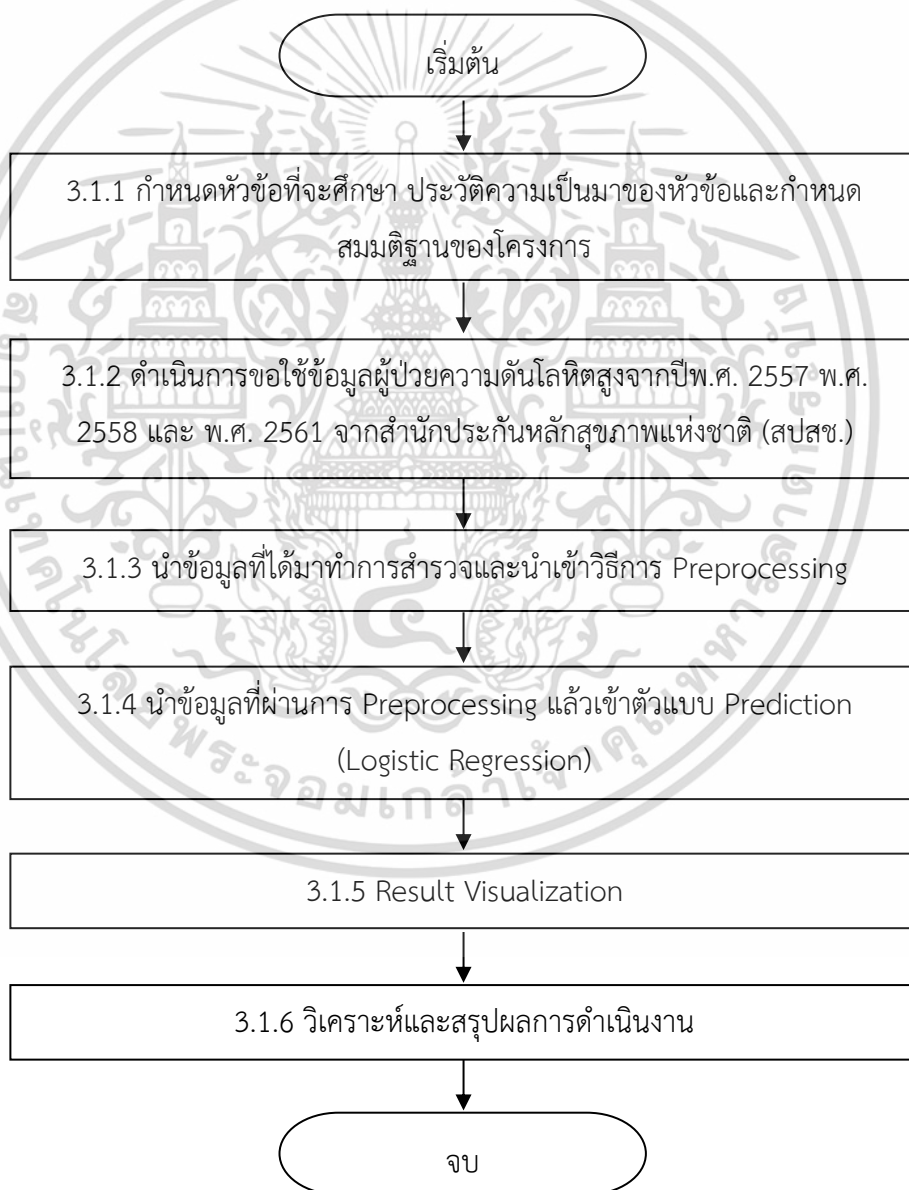
บทที่ 3

วิธีการดำเนินงาน

ในบทนี้จะกล่าวถึงระยะเวลาและวิธีการดำเนินงานซึ่งขั้นตอนทั้งหมดตั้งแต่เริ่มดำเนินการ ประกอบไปด้วยการใช้ Dataset การทำ Preprocessing การ Prediction และการทำ Visualization

3.1 ระเบียบวิธีการดำเนินงาน

ระเบียบวิธีการดำเนินงานนี้มีขั้นตอนแสดงดังรูปที่ 3.1



รูปที่ 3.1 ระเบียบวิธีการดำเนินงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.1 กำหนดหัวข้อ ความเป็นมาและความสำคัญของหัวข้อและกำหนดสมมติฐาน

กำหนดหัวข้อที่สนใจและศึกษางานวิจัยที่เกี่ยวข้องเพื่อค้นหาความเป็นมาและความสำคัญ เพื่อนำข้อมูลมาเรียบเรียงแล้วนำไปตั้งเป็นสมมติฐานในหัวข้อ “การคาดการณ์แนวโน้มความชุกและปัจจัยที่สัมพันธ์กับโรคอ้วนในผู้ป่วยความดันโลหิตสูงในประเทศไทย” ที่ผู้วิจัยสนใจจะศึกษา

3.1.2 ดำเนินการขอใช้ข้อมูลผู้ป่วยความดันโลหิตสูงจากปีพ.ศ. 2557 พ.ศ. 2558 และพ.ศ. 2561 จากสำนักประกันหลักสุขภาพแห่งชาติ (สปสช.)

ดำเนินการขอข้อมูลผู้ป่วยความดันโลหิตสูงจากปีพ.ศ. 2557 พ.ศ. 2558 และพ.ศ. 2561 จากสำนักประกันหลักสุขภาพแห่งชาติ (สปสช.) ที่อยู่ภายใต้การดูแลของกระทรวงสาธารณสุขซึ่งต้องยื่นขอข้อมูลสำหรับดำเนินงานผ่านโครงการ Data Archival for Maximum Utilization System (DAMUS) ขั้นตอนการยื่นเป็นไปตามหลักเกณฑ์มาตรฐานของ สปสช. โดยข้อมูลนั้นจะประกอบไปด้วยข้อมูลตาม Case record form ที่ถูกกำหนดโดย สปสช. ร่วมกับเครือข่ายวิจัยกลุ่มสถาบันแพทยศาสตร์แห่งประเทศไทย ซึ่งประกอบไปด้วยข้อมูลเช่น เพศ อายุ อาชีพ การนับถือศาสนา สัญชาติ สิทธิการรักษา การสูบบุหรี่ การดื่มสุรา โรคร่วมอื่นๆ ผลเลือดที่เคยได้รับการตรวจ เช่น FPG, HbA1c เป็นต้น

3.1.3 นำข้อมูลที่ได้มาทำการสำรวจและนำเข้าวิธีการ Preprocessing

เมื่อนำข้อมูลมาทำการสำรวจพบว่าในปีพ.ศ. 2557 มีข้อมูลทั้งหมด 260 Attributes 33,227 Records ปีพ.ศ. 2558 มีข้อมูลทั้งหมด 255 Attributes 32,440 Records ปีพ.ศ. 2561 มีข้อมูลทั้งหมด 353 Attributes 38,448 Records เนื่องจากมีข้อมูลที่เป็น Missing Value, Attribute ที่มีความเกี่ยวข้องน้อย และ Attribute ที่จำเป็นต้องแปลความเพิ่มเติมจำนวนมาก จึงต้องทำการ Preprocessing data ก่อนนำเข้า Prediction Model โดยแสดงรายละเอียดในหัวข้อที่ 3.3

3.1.4 นำข้อมูลที่ผ่านการ Preprocessing แล้วเข้าตัวแบบ Prediction (Logistic Regression)

เมื่อข้อมูลถูก Preprocessing เรียบร้อยแล้วจึงนำข้อมูลมาหาค่าสถิติเพื่อดู insight ของข้อมูลแล้วนำข้อมูลไปเข้าตัวแบบ Logistic Regression เพื่อคาดการณ์แนวโน้มความชุกและปัจจัยเสี่ยงของโรคอ้วนในผู้ป่วยความดันโลหิตสูงในประเทศไทย โดยแสดงรายละเอียดในหัวข้อที่ 3.4

3.1.5 Result Visualization

แสดงผลตาราง Evaluation Matrix (Accuracy, Precision, Recall, F1-score) และแสดงผลของ Insight จากข้อมูลในแง่มุมต่างๆ และนำเอาการคาดการณ์แนวโน้มโรคอ้วนในผู้ป่วย ความดันโลหิตสูงไปเทียบกับแนวโน้มโรคอ้วนในประเทศไทย โดยแสดงรายละเอียดในหัวข้อที่ 3.5

3.1.6 วิเคราะห์และสรุปผลการดำเนินงาน

ในขั้นตอนนี้จะนำเอาผลลัพธ์ทั้งหมดที่ได้มาวิเคราะห์ เพื่อสรุปเป็นองค์ความรู้ที่สามารถนำไปประยุกต์ใช้กับสาธารณสุขเพื่อเพิ่มประสิทธิภาพการให้บริการและส่งเสริมสุขภาพของผู้ป่วยให้เปลี่ยนแปลงไปในทางที่ดีขึ้นได้จริง

3.2 ระยะเวลาการดำเนินงาน

การดำเนินงานจะเริ่มในช่วงเดือนตุลาคม พ.ศ. 2564 และมาสิ้นสุดในช่วงเดือนกรกฎาคม พ.ศ. 2565 เป็นระยะเวลาประมาณ 10 เดือนโดยมีรายละเอียดตามตารางที่ 3.1

ตารางที่ 3.1 ตารางการดำเนินงาน

ขั้นตอน	การดำเนินงาน	ระยะเวลาดำเนินงาน				
		ต.ค. 64	พ.ย. 64	ธ.ค. 64 - เม.ย. 65	พ.ค. 65 - มิ.ย. 65	ก.ค. 65
1.	กำหนดหัวข้อที่จะศึกษา	X				
2.	ดำเนินการขอใช้ข้อมูล		X			
3.	การทำ Preprocessing			X		
4.	การนำข้อมูลที่ผ่านการ Preprocessing แล้วเข้า Model			X		
5.	Data Visualization			X		
6.	ประมวลผลข้อมูลและวิเคราะห์ผล				X	
7.	รายงานผลและนำเสนอ					X

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

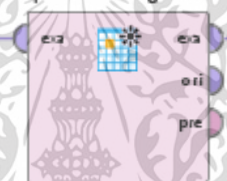
3.3 ขั้นตอนการ Preprocessing

ในขั้นตอนการ Preprocessing เบื้องต้นจะใช้โปรแกรม RapidMiner ในการดำเนินการ เนื่องจากตัวข้อมูลที่ได้มาจากการเก็บข้อมูลทุติยภูมิทำให้ Attribute ที่ได้มีจำนวนมาก ซึ่งโปรแกรม RapidMiner ที่เป็นโปรแกรมสำเร็จรูปจะช่วยให้เห็นโครงและรายละเอียดของข้อมูลได้สะดวกขึ้น เพื่อให้ง่ายต่อการ Cleansing โดยรายละเอียดขั้นตอนจะมีดังต่อไปนี้

3.3.1 Missing Value

ในวิธีการแรกจะใช้ Function “Replace Missing Values” ดังในรูปที่ 3.2 เพื่อ Cleansing Missing Value ออกจากข้อมูลโดยที่ข้อมูลอาทิเช่น ความสูง (Height) น้ำหนัก (Weight) รอบเอว (waist) จะใช้การ Replace with average และข้อมูลที่ไม่สามารถแทนที่ด้วยค่าตัวเลขได้อาทิเช่น ค่าความดันโลหิตต่ำที่สุด (SBP) ค่าความดันโลหิตสูงสุด (DBP) จะใช้การ Replace with zero

รูปที่ 3.2 Function “Replace Missing Values”



รูปที่ 3.2 Function “Replace Missing Values”

3.3.2 Generate Attribute

ในวิธีการที่ 2 จะใช้ Function “Generate Attribute” ดังในรูปที่ 3.3 โดยจะใช้ใน 2 ลักษณะคือ 1) เพื่อสร้าง Attribute ใหม่โดยคำนวณค่า value ใหม่จาก value ที่มีอยู่แล้ว 2) ตีความข้อมูลตัวเลขให้อยู่ในรูปแบบของข้อมูลตัวอักษรที่สื่อความได้มากกว่า ดังตารางที่ 3.2

รูปที่ 3.3 Function “Generate Attribute”



รูปที่ 3.3 Function “Generate Attribute”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 Function Expression ใน Function “Generate Attribute”

Attribute Name	Function Expression
bmi	(weight)/(height/100)^2
bmi_group	If (bmi < 18.5,"UnderWeight", If (bmi >= 18.5 && bmi < 23,"Normal", If (bmi >= 23 && bmi < 25,"OverWeight", "Obesity"))
dbp_count	If (dbp_1 <= 0 && dbp_2 <=0 && dbp_3 <= 0 , 0, If (dbp_1 <= 0 && dbp_2 <=0 , 1, If (dbp_2 <= 0 && dbp_3 <=0 , 1, If (dbp_1 <= 0 && dbp_3 <=0 , 1, If (dbp_1 <= 0 ,2, If (dbp_2 <= 0 ,2, If (dbp_3 <= 0 ,2, 3))))))
sbp_count	If (sbp_1 <= 0 && sbp_2 <=0 && sbp_3 <= 0 , 0, If (sbp_1 <= 0 && sbp_2 <=0 , 1, If (sbp_2 <= 0 && sbp_3 <=0 , 1, If (sbp_1 <= 0 && sbp_3 <=0 , 1, If (sbp_1 <= 0 ,2, If (sbp_2 <= 0 ,2, If (sbp_3 <= 0 ,2, 3))))))
dbp_avg	(dbp_1 + dbp_2 + dbp_3)/dbp_count
sbp_avg	(sbp_1 + sbp_2 + sbp_3)/sbp_count
smoking	If (smoking == 3 , "no" , "yes")
cholesterol	If (tc_res > 200 , "high" , "normal")
triglyceride	If (tg_res > 150 , "high" , "normal")
hdl	If (hdl_res <= 40 , "Low" , "normal")
ldl	If (ldl_res > 130 , "high" , "normal")
obesity	If (bmi >= 25 , "1" , "0")
occ_type	If (occ > 0 && occ <= 8 , "indoor" , "outdoor")

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

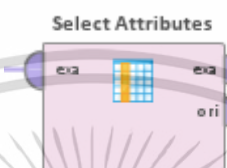
ตารางที่ 3.2 Function Expression ใน Function “Generate Attribute” (ต่อ)

Attribute Name	Function Expression
ckd_stage	If (EGFR < 15,"G5", If (EGFR >= 15 && EGFR <= 29,"G4", If (EGFR >= 30 && EGFR <= 44,"G3b", If (EGFR >= 45 && EGFR <= 59,"G3a", If (EGFR >= 60 && EGFR <= 89,"G2", "G1"))))
age_group	If (age < 30,"Young Adult", If (age >= 30 && age < 50,"Adult", If (age >= 50 && age < 70,"Senior", "Elder"))
area_group	If (area < 4,"North", If (area >= 4 && area < 7,"Center", If (area >= 7 && area < 11,"North-East", If (area >= 11 && area < 13,"South", "Bangkok"))))
Metabolic_synd	If (gender == male , (waist >= 80 , “yes”, ”no”) , (waist >=90, “yes”, ”no”))
diabetes	If (fpg1_res >= 126,"yes", "no")

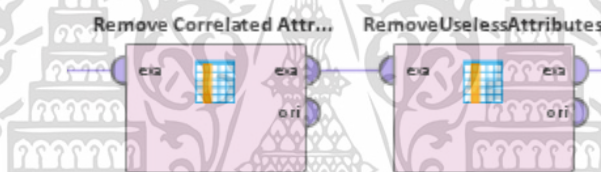
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.3 Select Attribute

ในวิธีการที่ 3 จะใช้ Function “Select Attribute” ดังในรูปที่ 3.4 เพื่อเลือกเอาเฉพาะ Attribute ที่จำเป็นโดยคัดกรองจากขั้นที่ 1) Delete Attribute ที่มี Missing value มากกว่า 10000 records ขั้นที่ 2) ใช้ function remove correlated attribute เพื่อตัด attribute ที่สัมพันธ์กันออก และใช้ function remove useless attribute ดังในรูปที่ 3.5 เพื่อตัด attribute ที่ไม่ส่งผลกับตัวแปรตามออก และขั้นที่ 3) Review จากแพทย์ผู้ซึ่งมีความรู้โดยตรงจากโรคความดันโลหิตสูง โดยจะได้ผลลัพธ์ตามตารางที่ 3.3



รูปที่ 3.4 Function “Select Attribute”



รูปที่ 3.5 Function “remove attribute”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.3 Attributes ที่จะนำไปใช้ต่อในขั้นตอนถัดไป

Attribute Name	Attribute Detail
obesity	กำหนดให้เป็นค่าตัวแปรตาม ใช้สำหรับระบุเกณฑ์โรคอ้วนในผู้ป่วยความดันโลหิตสูง โดยกำหนดเป็น bmi มากกว่าหรือเท่ากับ 25 kg/m ² เป็นโรคอ้วนแทนด้วย “1” จำนวน 34,600 records, bmi น้อยกว่า 25 kg/m ² ไม่เป็นโรคอ้วนแทนด้วย “0” จำนวน 40,574 records
age	ระบุอายุของผู้ป่วยความดันโลหิตสูง
gender	ระบุเพศของผู้ป่วยความดันโลหิตสูง
weight	ระบุน้ำหนักของผู้ป่วยความดันโลหิตสูง
waist	ระบุรอบเอวของผู้ป่วยความดันโลหิตสูง
zone	ระบุตำแหน่งของสถานที่การรักษาจาก 13 เขตสุขภาพของประเทศไทย
dbp_avg	ค่าเฉลี่ยความดันโลหิตต่ำสุด (DBP : Diastolic Blood Pressure) จากทุกครั้งของการเข้ารับการรักษา
sbp_avg	ค่าเฉลี่ยความดันโลหิตสูงที่สุด (SBP : Systolic Blood Pressure) จากทุกครั้งของการเข้ารับการรักษา
smoking	ระบุการสูบบุหรี่ของผู้ป่วยความดันสูง โดยกำหนดเป็นไม่สูบแทนด้วย “no”, สูบแทนด้วย “yes”
cholesterol	ไขมันประเภทเลวชนิดหนึ่ง ค่ามากกว่า 200 แทนด้วย “high”, ค่าน้อยกว่าหรือเท่ากับ 200 แทนด้วย “normal”
triglyceride	ไขมันประเภทเลวชนิดหนึ่ง ค่ามากกว่า 150 แทนด้วย “high”, ค่าน้อยกว่าหรือเท่ากับ 150 แทนด้วย “normal”
hdl	ไขมันประเภทดีชนิดหนึ่ง ค่าน้อยกว่าหรือเท่ากับ 40 แทนด้วย “low”, ค่ามากกว่า 40 แทนด้วย “normal”
ldl	ไขมันประเภทเลวชนิดหนึ่ง ค่ามากกว่า 130 แทนด้วย “high”, ค่าน้อยกว่าหรือเท่ากับ 130 แทนด้วย “normal”
occ_type	ระบุอาชีพของผู้ป่วยความดันโลหิตสูง โดยกำหนดเป็นการทำงานแบบ “indoor” และ “outdoor”
ckd_stage	ระยะของโรคไตแบบเรื้อรังจากอัตราการกรองของเสียของไต (EGFR : Estimated glomerular filtration rate) โดยกำหนดเป็น 6 ชั้นได้แก่ G1 : ปกติ, G2, G3a, G3b, G4, G5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.3 Attributes ที่จะนำไปใช้ต่อในขั้นตอนถัดไป (ต่อ)

Attribute Name	Attribute Detail
diabetes	ระบุโรคเบาหวานของผู้ป่วยความดันโลหิตสูงจากปริมาณน้ำตาลในเลือด (FPG : fasting plasma glucose) หากเป็นโรคเบาหวานแทนด้วย “1”, ไม่เป็นเบาหวานแทนด้วย “0”
age_group	กำหนดช่วงอายุเป็นทั้งหมด 4 ช่วง อายุต่ำกว่า 30 เป็น Young Adult อายุตั้งแต่ 30 ปีถึง 49 ปีเป็น Adult อายุตั้งแต่ 50 ปีถึง 69 ปีเป็น Senior และอายุมากกว่า 70 ปีขึ้นไปเป็น Elder
area_group	จัดกลุ่มของ 13 เขตสุขภาพในประเทศไทยแบ่งโดยตามภูมิภาค ภาคเหนือ ภาคตะวันออกเฉียงเหนือ ภาคกลาง ภาคใต้ และกรุงเทพมหานคร
metabolic_synd	ระบุผู้ป่วยโรคความดันโลหิตสูงที่เป็นโรคอ้วนลงพุง (Abdominal Obesity) โดยกำหนดให้ในเพศหญิงต้องมีขนาดรอบเอวมากกว่า 90 เซนติเมตร และเพศชายมากกว่า 80 เซนติเมตร

3.4 ขั้นตอนการ Preprocessing ด้วยภาษา Python

หลังจากการ Preprocessing เบื้องต้นจากโปรแกรม RapidMiner แล้วจะมีขั้นตอนการเตรียมข้อมูลก่อนเข้าตัวแบบเพิ่มเติมดังนี้

1) การ Import Library

ทำการ Import Library ทุกตัวที่ต้องการใช้งานเช่น 1) Pandas ที่ช่วยในการวิเคราะห์ข้อมูล และการทำ data cleaning 2) Numpy ช่วยในเรื่องการคำนวณทางคณิตศาสตร์ จัดการกับ array หลายมิติ และการทำ Matrix 3) sklearn ช่วยในการสร้างตัวแบบเป็นต้น ดังโค้ดด้านล่าง

```
#import Library
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn import preprocessing
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from matplotlib import pyplot as plt
import seaborn as sns
import statsmodels.api as sm
from sklearn.decomposition import PCA
from sklearn.metrics import mean_squared_error, roc_curve, roc_auc_score
import math
from sklearn.model_selection import train_test_split
```

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์และสงวนสิทธิ์ในเนื้อหาและข้อมูลทั้งหมดไว้เป็นของตนเองและใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) การ Read File

ทำการอ่านไฟล์ dataset ข้อมูลของผู้ป่วยโรคความดันโลหิตสูงทั้ง 3 ปีเข้าไปเก็บไว้ในตัวแปร df2 ดังโค้ดด้านล่าง

```
#read file
df2 = pd.read_csv("/content/factor_vars.csv")
```

3) การสร้าง Dummy Variable

ทำการจัดการ Attributes ที่เป็น Category ให้เป็นค่า 0,1 เพื่อที่จะนำเข้าไป Logistic regression model ดังโค้ดด้านล่าง

```
#create dummy var
training=pd.get_dummies(df2, columns=["gender","smoking","cholesterol","triglyceride","hdl","ldl","occ_type","ckd_stage","diabetes","age_group","area_group","metabolic_synd"])

final_data = training
final_data.head()
```

4) การตรวจสอบรายละเอียดของข้อมูลและค่าสถิติเบื้องต้น

ทำการตรวจสอบค่า missing value, data type และตรวจสอบค่าสถิติเบื้องต้นอาทิเช่นค่า min, max, avg, standard deviation เป็นต้น ดังโค้ดด้านล่าง

```
#data information
final_data.info()
#data stat
final_data.describe().transpose()
```

5) การหาความสัมพันธ์กันในตัวแปรอิสระ

ทำการตรวจสอบหาตัวแปรอิสระที่สัมพันธ์กันเองเพื่อตัดออก โดยใช้ heatmap เข้ามาช่วยในการแสดงผล ดังโค้ดด้านล่าง

```
#X correlation
plt.subplots(figsize=(30, 15))
sns.heatmap(final_data.corr(), annot=True)
plt.show()
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6) การกำหนดตัวแปรอิสระและตัวแปรตาม

กำหนดชุดของตัวแปรอิสระที่จะใช้ในการเข้าตัวแบบทั้งหมดเก็บไว้ในตัวแปร X และเก็บ Attribute ที่จะเป็นตัวแปรตาม ไว้ในตัวแปร y ดังโค้ดด้านล่าง

```
#split dataset in features and target variable

feature_cols = ['age_group_Adult','age_group_Elder','age_group_Senior', 'age_group_Young Adult', 'area_group_Bangkok','area_group_Center', 'area_group_North', 'area_group_North-East', 'area_group_South','sbp_avg','dbp_avg','gender_male','smoking_no','smoking_yes','cholesterol_high','cholesterol_normal','triglyceride_high','triglyceride_normal','hdl_Low','hdl_normal',
'occ_type_indoor','occ_type_outdoor','ckd_stage_G1','ckd_stage_G2','ckd_stage_G3a','ckd_stage_G3b','ckd_stage_G4','ckd_stage_G5','diabetes_no','diabetes_yes','waist']

X = final_data[feature_cols] # Features
y = final_data.obesity # Target variable
```

7) การแบ่ง Train dataset และ Test dataset

ทำการแบ่ง Train dataset และ Test dataset ออกเป็นอัตราส่วน 7:3 แล้วเก็บในตัวแปร X_train,X_test,y_train,y_test ดังโค้ดด้านล่าง

```
# split X and y into training and testing sets
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.3,random_state=0)
```

8) การทำ PCA

ทำ Principal Component Analysis (PCA) เพื่อช่วยในการทำ decomposition ข้อมูล เพื่อลด features redundancy และช่วยบีบอัด feature space ให้เล็กลง ดังโค้ดด้านล่าง

```
#PCA
pca = PCA()
X_train = pca.fit_transform(X_train)
X_test = pca.transform(X_test)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5 การนำข้อมูลเข้าตัวแบบ Logistic Regression

เมื่อเตรียมข้อมูลจากขั้นตอนการ Preprocessing เรียบร้อยแล้วก็นำข้อมูลที่ได้นำเข้า Logistic Regression model เพื่อหาความสัมพันธ์ ดังโค้ดด้านล่าง

```
#instantiate the model (using the default parameters)
logreg = LogisticRegression()
#fit the model with data
logreg.fit(X_train,y_train)
#Logistic regression
y_pred=logreg.predict(X_test)
```

3.6 การแสดงผล Visualization

เมื่อได้ผลจากการนำข้อมูลที่ผ่านการ Preprocessing และเข้าตัวแบบเรียบร้อยแล้วจึงผลลัพธ์มาแสดงผลเป็นกราฟรูปแบบต่างๆให้เข้าใจได้ง่ายขึ้น

1) แสดงตาราง Confusion Matrix

ตาราง Confusion Matrix จะแสดงค่า True Positive, True Negative, False Positive และ False Negative ของตัวแบบดังโค้ดด้านล่าง

```
#Confusion Matrix
cm=confusion_matrix(y_test,y_pred)
print(cm)
plt.figure(figsize=(16,8))
plt.title("Confusion Matrix")
sns.heatmap(cm, annot=True,fmt='d', cmap='Blues')
plt.ylabel("Actual Values")
plt.xlabel("Predicted Values")
plt.savefig('confusion_matrix.png')
```

2) แสดงค่า Accuracy, Precision, Recall และ F1-score

เมื่อได้ค่า True Positive, True Negative, False Positive และ False Negative ของตัวแบบมาแล้วขั้นตอนนี้จะเป็นการนำชุดตัวเลขเข้าสู่สูตรเพื่อหาค่า Accuracy, Precision, Recall และ F1-score เพื่อบ่งบอกถึงประสิทธิภาพของตัวแบบดังกล่าว

```
#Accuracy, Precision, Recall, f1-score
print(classification_report(y_test,y_pred))
model_acc = logreg.score(X_train, y_train)
print("The Accuracy for Model is {}".format(model_acc*100))
```

3) แสดงค่า Mean Square Error และ Root Mean Square Error

แสดงค่า Mean Square Error เพื่อหาค่า Loss ของ y_{actual} และ y_{predic} โดยการนำผลต่างมายกกำลังสองหาค่าเฉลี่ย และค่า Root Mean Square Error ก็คือนำเอา MSE มาใส่ SQRT เนื่องจากตอนที่นำค่าผลต่างมายกกำลังสองเพื่อให้เครื่องหมายลบหายไป ทำให้หน่วยอาจจะเปลี่ยนไปจากเดิม

```
#MSE, RMSE
y_actual = y
y_predic = y_pred

MSE = mean_squared_error(y_actual, y_predic)

RMSE = math.sqrt(MSE)

print("Mean Square Error:")
print(MSE)
print("\n Root Mean Square Error:")
print(RMSE)
```

4) แสดงจำนวน records เมื่อเทียบกับ factor ต่างๆ

หาจำนวนของ records ที่เป็นโรคอ้วนและไม่เป็นโรคอ้วนใน factor ต่างๆ แล้ว Plot เป็นกราฟ ดังโค้ดด้านล่าง

```
#crosstab data
genderplot = pd.crosstab(df2['gender'], df2['obesity'] )
print (genderplot)
barplot = genderplot.plot.bar(rot=0)

smokingplot = pd.crosstab(df2['smoking'], df2['obesity'] )
print (smokingplot)
barplot = smokingplot.plot.bar(rot=0)

cholplot = pd.crosstab(df2['cholesterol'], df2['obesity'] )
print (cholplot)
barplot = cholplot.plot.bar(rot=0)

hdlplot = pd.crosstab(df2['hdl'], df2['obesity'] )
print (hdlplot)
barplot = hdlplot.plot.bar(rot=0)

occplot = pd.crosstab(df2['occ_type'], df2['obesity'] )
print (occplot)
barplot = occplot.plot.bar(rot=0)

ckdplot = pd.crosstab(df2['ckd_stage'], df2['obesity'] )
print (ckdplot)
barplot = ckdplot.plot.bar(rot=0)

diabepplot = pd.crosstab(df2['diabetes'], df2['obesity'] )
print (diabepplot)
barplot = diabepplot.plot.bar(rot=0)

areaplot = pd.crosstab(df2['area_group'], df2['obesity'])
print (areaplot)
barplot = areaplot.plot.bar(rot=0)

ageplot = pd.crosstab(df2['age_group'], df2['obesity'])
print (ageplot)
barplot = ageplot.plot.bar(rot=0)

metabplot = pd.crosstab(df2['metabolic_synd'], df2['obesity'])
print (metabplot)
barplot = metabplot.plot.bar(rot=0)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้拿去ใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) แสดงค่า P-value และ Odds ratio

เป็นขั้นตอนการทำ Multivariable factors เพื่อค่า P-value เป็นการแสดงความสัมพันธ์ของตัวแปรอิสระมีนัยสำคัญทางสถิติกับตัวแปรตามหรือไม่ โดยที่ตั้งค่า P-value ไว้ที่ < 0.005 และนำเอาค่า coefficient ของแต่ละตัวแปรอิสระไปหา Odds ratio ด้วยวิธีการ exponential ที่ค่า Confident Interval = 95% (0.95) ค่า Lower CI และค่า Upper CI ดังโค้ดด้านล่าง

```
#P-value,Odds ratio
reg = smf.logit("obesity ~ age_group+metabolic_synd+area_group+
C(gender, Treatment('male'))+C(smoking,Treatment('yes'))+C(chol
esterol,Treatment('normal')) +C(hdl, Treatment('normal'))+occ_t
ype+ckd_stage+diabetes" ,data=df2).fit()
print(reg.summary(alpha=0.1))
params = reg.params
conf = reg.conf_int(alpha=0.1)
conf['OR'] = params
conf.columns = ["Lower CI", "Upper CI", "OR"]
np.exp(conf)
```

6) แสดงกราฟ ROC Curve

เป็นขั้นตอนในการหา Test Performance สำหรับตัวแบบเพื่อป้องกัน Test dataset นั้นสามารถจำแนกกลุ่มโรคออกได้ดีเพียงใด ดังโค้ดด้านล่าง

```
# plot ROC Curve
fpr, tpr, thresholds = roc_curve(y_test, y_pred, pos_label = 1)
plt.figure(figsize=(6,4))
plt.plot(fpr, tpr, linewidth=2)
plt.plot([0,1], [0,1], 'k--')
plt.rcParams['font.size'] = 12
plt.title('ROC curve for Obesity classifier')
plt.xlabel('False Positive Rate (1 - Specificity)')
plt.ylabel('True Positive Rate (Sensitivity)')
plt.show()
# Area
auc = roc_auc_score(y_test, y_pred)
print (auc)
```

บทที่ 4

ผลการดำเนินงาน

ในบทนี้จะกล่าวถึงผลของการดำเนินงาน การแสดงข้อมูลของกลุ่มตัวอย่าง รวมไปถึงผลลัพธ์ หลังการนำข้อมูลเข้าตัวแบบที่อยู่ในรูปแบบของตารางและกราฟรูปแบบต่างๆ พร้อมทั้งคำอธิบาย

4.1 ลักษณะของกลุ่มตัวอย่าง

ลักษณะของกลุ่มตัวอย่างเป็นผู้ป่วยโรคความดันโลหิตสูงในประเทศไทยทั้งหมด 3 ปีรวม 79663 records ในปี 2557 มีกลุ่มตัวอย่าง 25647 records เป็นเพศชาย 9556 records (ร้อยละ 37.3) เป็นเพศหญิง 16091 records (ร้อยละ 62.7) อายุเฉลี่ยของกลุ่มตัวอย่างคือ 63.8 ± 11.6 ปี โดยส่วนใหญ่ทำงานกลางแจ้งร้อยละ 55.18 และมีรอบเวาเฉลี่ยของกลุ่มตัวอย่างคือ 61.8 ± 12.9 นีว ในปี 2558 มีกลุ่มตัวอย่าง 21795 records เป็นเพศชาย 8244 records (ร้อยละ 37.8) เป็นเพศหญิง 13551 records (ร้อยละ 62.2) อายุเฉลี่ยของกลุ่มตัวอย่างคือ 63.9 ± 11.7 ปี โดยส่วนใหญ่ทำงานกลางแจ้งร้อยละ 55.04 และมีรอบเวาเฉลี่ยของกลุ่มตัวอย่างคือ 62.3 ± 13.2 นีว และในปี 2561 มีกลุ่มตัวอย่าง 32221 records เป็นเพศชาย 12219 records (ร้อยละ 37.9) เป็นเพศหญิง 20002 records (ร้อยละ 62.1) อายุเฉลี่ยของกลุ่มตัวอย่างคือ 64.6 ± 11.8 ปี โดยส่วนใหญ่ทำงานกลางแจ้งร้อยละ 53.28 และมีรอบเวาเฉลี่ยของกลุ่มตัวอย่างคือ 62.5 ± 13.2 นีว ดังตารางที่ 4.1

ตารางที่ 4.1 ข้อมูลลักษณะทั่วไปของกลุ่มตัวอย่าง

Characteristics	2557 n = 25647 n (%)	2558 n = 21795 n (%)	2561 n = 32221 n (%)
Gender			
▪ Male	9556 (37.3)	8244 (37.8)	12219 (37.9)
▪ Female	16091 (62.7)	13551 (62.2)	20002 (62.1)
Age (year) mean \pm SD	63.8 ± 11.6	63.9 ± 11.7	64.6 ± 11.8
▪ 18 - 20	3 (0.01)	1 (0.005)	0 (0)
▪ 20 - 29	31 (0.12)	29 (0.13)	46 (0.14)
▪ 30 - 39	392 (1.53)	381 (1.75)	453 (1.40)
▪ 40 - 49	2498 (9.74)	2023 (9.28)	2776 (8.62)
▪ 50 - 59	6402 (24.96)	5514 (25.30)	7637 (23.70)
▪ 60 - 69	7926 (30.90)	6589 (30.23)	10237 (31.77)
▪ 70 - 79	6042 (23.56)	5179 (23.76)	7537 (23.39)
▪ \geq 80	2353 (9.17)	2079 (9.54)	3535 (10.97)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

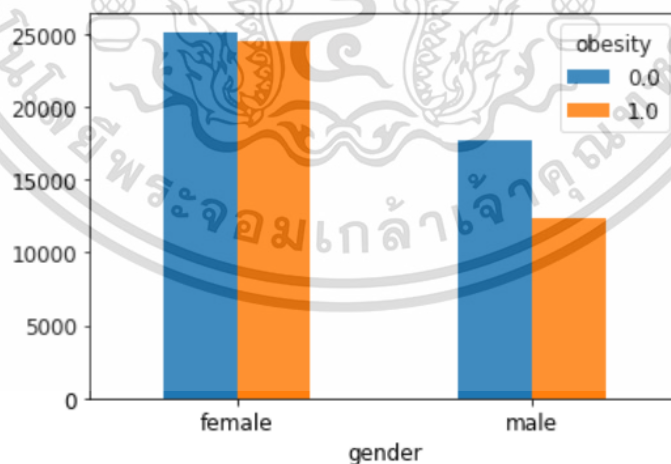
ตารางที่ 4.1 ข้อมูลลักษณะทั่วไปของกลุ่มตัวอย่างต่อ

Characteristics	2557 n = 25647 n (%)	2558 n = 21795 n (%)	2561 n = 32221 n (%)
Occupation			
▪ Indoor	11494 (44.82)	9800 (44.96)	15053 (46.72)
▪ Outdoor	14153 (55.18)	11995 (55.04)	17168 (53.28)
Waist (cm)	61.8 ± 12.9	62.3 ± 13.2	62.5 ± 13.2
▪ < 30	14 (0.054)	0 (0)	152 (0.47)
▪ 30 - 59	60 (0.23)	91 (0.42)	990 (3.07)
▪ 60 - 89	19037 (74.227)	15307 (70.23)	20539 (63.74)
▪ 90 - 119	6448 (25.14)	6308 (28.94)	10377 (32.21)
▪ ≥ 120	82 (0.32)	78 (0.36)	158 (0.49)

4.2 จำนวนประชากรแบ่งแยกตามการเป็นโรคอ้วน

4.2.1 Gender Group

จำนวนประชากรผู้ป่วยความดันโลหิตสูงที่แบ่งแยกตามการเป็นโรคอ้วนในกลุ่มเพศหญิงและเพศชายมีความแตกต่างที่ชัดเจนและมีจำนวนผู้ป่วยที่ไม่เป็นโรคอ้วนมากกว่าจำนวนผู้ป่วยที่เป็นโรคอ้วน โดยในผู้หญิงพบผู้ป่วยที่เป็นโรคอ้วน 49.3% และไม่เป็นโรคอ้วน 50.7% ขณะที่ในเพศชายพบผู้ป่วยที่เป็นโรคอ้วน 41.0% และไม่เป็นโรคอ้วน 59.0% ดังที่แสดงในรูปที่ 4.1

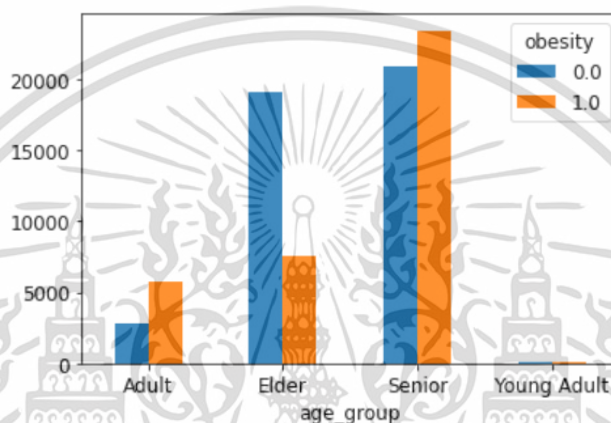


รูปที่ 4.1 Gender Group

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 Age Group

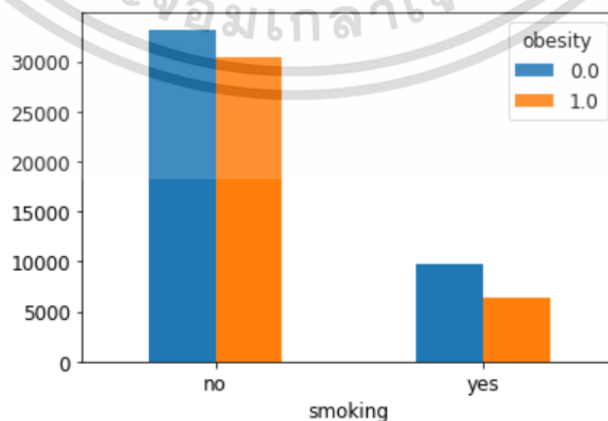
จำนวนประชากรผู้ป่วยความดันโลหิตสูงที่แบ่งแยกตามการเป็นโรคอ้วนที่แบ่งกลุ่มตามช่วงอายุโดยในช่วงวัย Young Adult (18-29 ปี) ที่มีจำนวนผู้ป่วยน้อยที่สุดในทุกกลุ่มอย่างเห็นได้ชัดพบผู้ป่วยที่เป็นโรคอ้วน 68.2% และไม่เป็นโรคอ้วน 31.8% ลำดับถัดมาเป็นช่วงวัย Adult (30-49 ปี) พบผู้ป่วยที่เป็นโรคอ้วน 66.9% และไม่เป็นโรคอ้วน 33.1% ลำดับถัดมาในช่วงวัย Senior (50-69 ปี) เป็นกลุ่มที่มีจำนวนผู้ป่วยมากที่สุดพบผู้ป่วยที่เป็นโรคอ้วน 52.9% และไม่เป็นโรคอ้วน 47.1% ลำดับถัดมาในช่วงวัย Elder (>70 ปี) พบผู้ป่วยที่เป็นโรคอ้วน 28.4% และไม่เป็นโรคอ้วน 71.6% ดังที่แสดงในรูปที่ 4.2



รูปที่ 4.2 Age Group

4.2.3 Smoking

จำนวนประชากรผู้ป่วยความดันโลหิตสูงที่แบ่งแยกตามพฤติกรรมการสูบบุหรี่ที่มีความแตกต่างกันค่อนข้างชัดเจนระหว่าง 2 กลุ่ม โดยพบผู้ป่วยที่เป็นโรคอ้วนในกลุ่มผู้ป่วยที่มีประวัติสูบบุหรี่ 39.7% และไม่เป็นโรคอ้วน 60.3% ในอีกทางหนึ่งพบผู้ป่วยที่เป็นโรคอ้วนแต่ไม่มีประวัติสูบบุหรี่ 47.8% และไม่เป็นโรคอ้วน 60.3% ดังแสดงในรูปภาพที่ 4.3

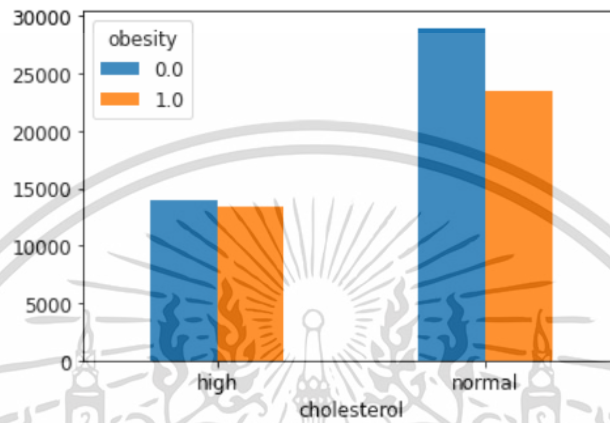


รูปที่ 4.3 Smoking

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.4 Cholesterol (bad fats)

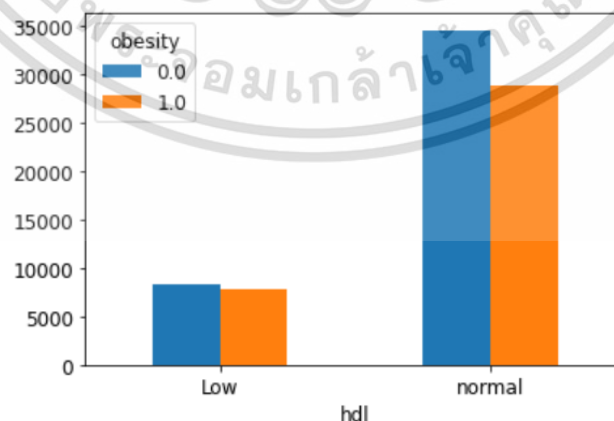
จำนวนประชากรผู้ป่วยโรคความดันโลหิตสูงที่แบ่งแยกตามค่าของคลอเลสเตอรอล (ไขมันไม่ดี) มีขนาดประชากรต่างกันประมาณกึ่งหนึ่ง โดยในผู้ป่วยที่มีคลอเลสเตอรอลในเลือดสูงจะพบผู้ป่วยที่เป็นโรคอ้วน 48.9% และไม่เป็นโรคอ้วน 51.1% โดยที่ผู้ป่วยที่มีคลอเลสเตอรอลในเลือดปกติจะพบผู้ป่วยที่เป็นโรคอ้วน 44.8% และไม่เป็นโรคอ้วน 55.2% ดังแสดงในรูปที่ 4.4



รูปที่ 4.4 Cholesterol (bad fats)

4.2.5 HDL (good fats)

จำนวนประชากรผู้ป่วยโรคความดันโลหิตสูงที่แบ่งแยกตามค่าของ HDL (ไขมันดี) มีขนาดประชากรต่างกันค่อนข้างมาก โดยในผู้ป่วยที่มี HDL ในเลือดต่ำจะพบผู้ป่วยที่เป็นโรคอ้วน 48.7% และไม่เป็นโรคอ้วน 51.3% ในขณะที่ผู้ป่วยที่มีค่า HDL ในเลือดปกติจะพบผู้ป่วยที่เป็นโรคอ้วน 54.4% และไม่เป็นโรคอ้วน 45.6% ดังแสดงในรูปที่ 4.5

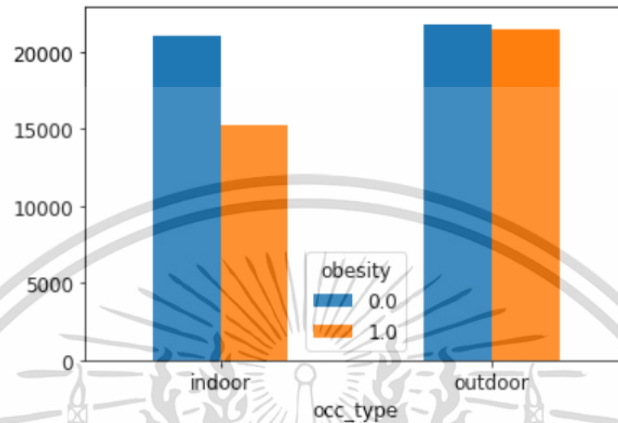


รูปที่ 4.5 HDL (good fats)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.6 Occupation Type

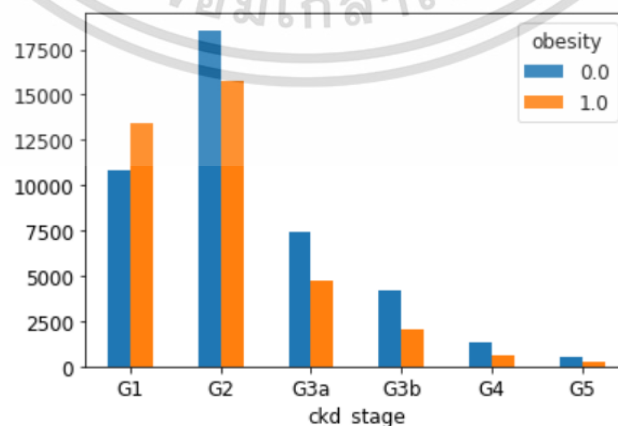
จำนวนประชากรผู้ป่วยโรคความดันโลหิตสูงแบ่งแยกตามการทำงานในร่มและทำงานกลางแจ้งพบว่าจำนวนประชากรทั้งสองกลุ่มมีความใกล้เคียงกันโดยผู้ป่วยที่ทำงานในร่มพบผู้ป่วยที่เป็นโรคอ้วน 42.1% และไม่เป็นโรคอ้วน 57.9% ในขณะที่ผู้ป่วยที่ทำงานกลางแจ้งพบผู้ป่วยที่เป็นโรคอ้วน 49.7% และไม่เป็นโรคอ้วน 50.3% ดังแสดงในรูปที่ 4.6



รูปที่ 4.6 Occupation Type

4.2.7 Chronic Kidney Disease Stages

จำนวนประชากรผู้ป่วยโรคความดันโลหิตสูงแบ่งแยกตามระยะของโรคไตเรื้อรังทั้งหมด 6 กลุ่ม (G1-G5) โดยพบว่ากลุ่มผู้ป่วยที่มีโรคไตเรื้อรังระยะ G2 มีจำนวนมากที่สุดและเป็นโรคอ้วน 45.9% ไม่เป็นโรคอ้วน 54.1% ลำดับถัดมาคือระยะ G1 พบผู้ป่วยที่เป็นโรคอ้วน 55.3% และไม่เป็นโรคอ้วน 44.7% ลำดับถัดมาคือระยะ G3a พบผู้ป่วยที่เป็นโรคอ้วน 38.6% และไม่เป็นโรคอ้วน 61.4% ลำดับถัดมาคือระยะ G3b พบผู้ป่วยที่เป็นโรคอ้วน 33.3% และไม่เป็นโรคอ้วน 66.7% ลำดับถัดมาคือระยะ G4 พบผู้ป่วยที่เป็นโรคอ้วน 31.5% และไม่เป็นโรคอ้วน 68.5% และระยะสุดท้าย G5 พบผู้ป่วยที่เป็นโรคอ้วน 33.9% และไม่เป็นโรคอ้วน 66.1% ดังแสดงในรูป 4.7

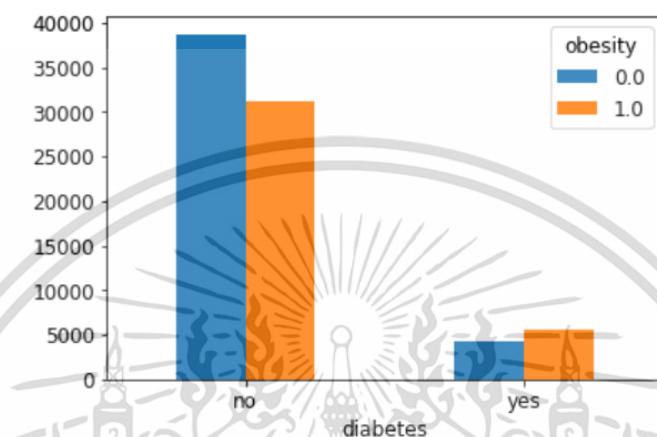


รูปที่ 4.7 Chronic Kidney Disease Stages

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.8 Diabetes

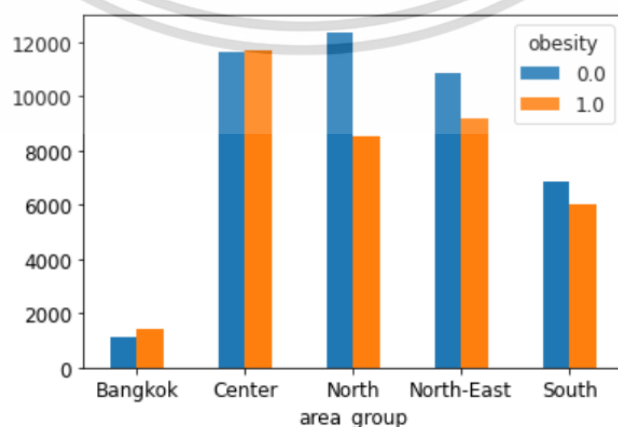
จำนวนประชากรผู้ป่วยโรคความดันโลหิตสูงแบ่งแยกตามการเป็นและไม่เป็นโรคเบาหวาน พบว่าจำนวนกลุ่มประชากรค่อนข้างมีขนาดต่างกันมาก โดยผู้ป่วยที่มีอาการแทรกซ้อนของโรคเบาหวานพบว่าเป็นโรคอ้วน 57.3% และไม่เป็นโรคอ้วน 42.7% ในขณะที่ผู้ป่วยที่ไม่มีอาการแทรกซ้อนของโรคเบาหวานพบว่าเป็นโรคอ้วน 44.6% และไม่เป็นโรคอ้วน 55.4% ดังแสดงในรูปที่ 4.8



รูปที่ 4.8 Diabetes

4.2.9 Area Group

จำนวนประชากรผู้ป่วยโรคความดันโลหิตสูงแบ่งแยกตามภูมิภาคของประเทศไทยพบว่าจำนวนประชากรของจังหวัดกรุงเทพมหานครแตกต่างจากภาคอื่นค่อนข้างมากเนื่องจากมีเพียงจังหวัดเดียว โดยผู้ป่วยในกรุงเทพมหานครพบว่าเป็นโรคอ้วน 55.5% และไม่เป็นโรคอ้วน 44.5% ถัดมาเป็นผู้ป่วยในภาคกลางพบว่าเป็นโรคอ้วนและไม่เป็นโรคอ้วนในอัตราส่วนใกล้เคียงกันประมาณ 50.0% ถัดมาเป็นผู้ป่วยภาคเหนือพบว่าเป็นโรคอ้วน 40.8% และไม่เป็นโรคอ้วน 59.2% พื้นที่ใกล้เคียงกันอย่างภาคตะวันออกเฉียงเหนือพบว่าเป็นโรคอ้วน 45.8% และไม่เป็นโรคอ้วน 54.2% และภาคใต้พบว่าเป็นโรคอ้วน 46.9% และไม่เป็นโรคอ้วน 53.1% ดังแสดงในรูปที่ 4.9

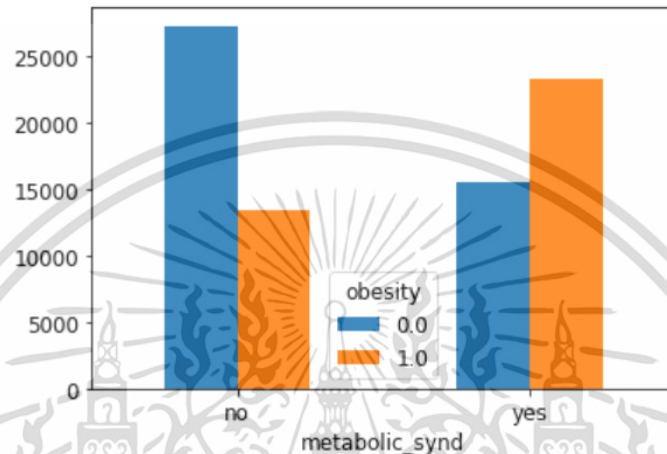


รูปที่ 4.9 Area Group

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.10 Abdominal Obesity

จำนวนประชากรผู้ป่วยโรคความดันโลหิตสูงแบ่งแยกตามการเป็นและไม่เป็นโรคอ้วนลงพุง พบว่าจำนวนกลุ่มประชากรใกล้เคียงกัน โดยผู้ป่วยที่มีอาการแทรกซ้อนของโรคอ้วนลงพุงพบว่าเป็นโรคอ้วน 60.0% และไม่เป็นโรคอ้วน 40.0% ในขณะที่ผู้ป่วยที่ไม่มีอาการแทรกซ้อนของโรคอ้วนลงพุงพบว่าเป็นโรคอ้วน 33.0% และไม่เป็นโรคอ้วน 67.0% ดังแสดงในรูปที่ 4.10

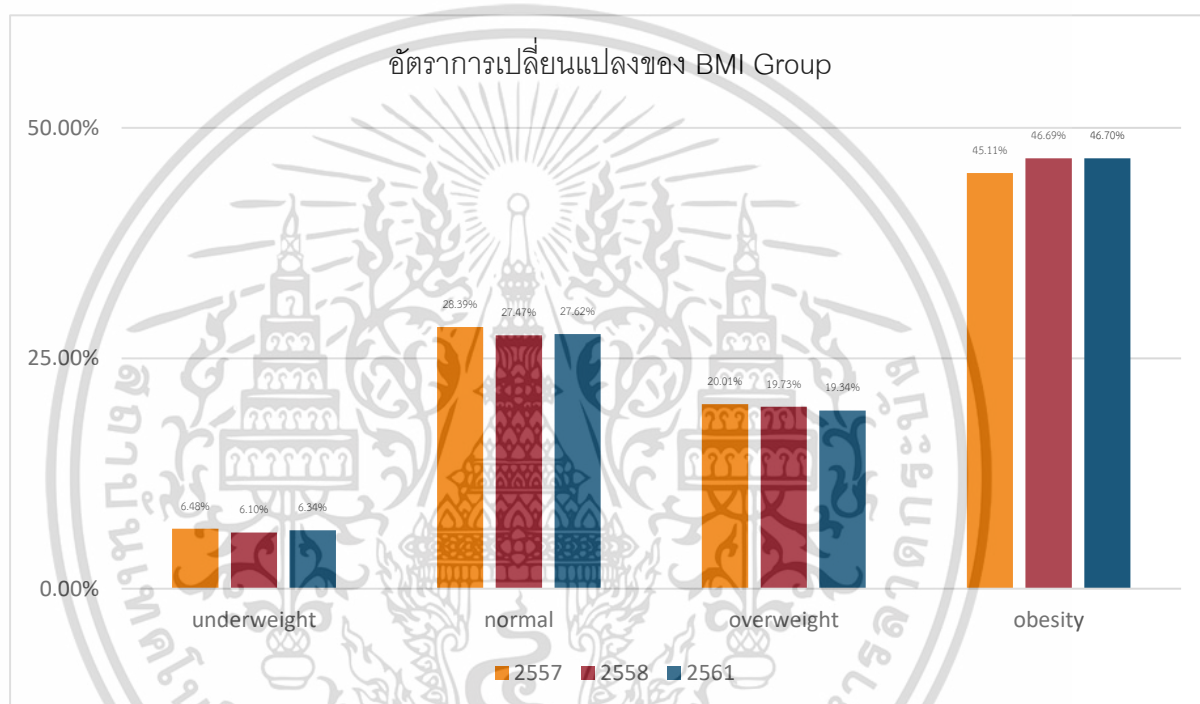


รูปที่ 4.10 Abdominal Obesity

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 อัตราการเปลี่ยนแปลงของ BMI Group ในแต่ละปี

ค่า BMI (Body Mass Index) ในกลุ่ม Asia-Pacific (Thailand) จะถูกจัดตามกลุ่มช่วงดังนี้ กลุ่มช่วง Underweight จะมีค่า BMI อยู่ในช่วงน้อยกว่า 18.5 kg/m² กลุ่มช่วง Normal จะอยู่ในช่วง 18.5 – 22.9 kg/m² กลุ่มช่วง Overweight จะอยู่ในช่วง 23 – 24.9 kg/m² และกลุ่มช่วง Obesity หรือเป็นโรคอ้วนจะอยู่ในช่วงตั้งแต่ 25 kg/m² เป็นต้นไป โดยเมื่อนำข้อมูลค่า BMI ของแต่ละช่วงมาเปรียบเทียบโดยปีจะพบว่าในทุกช่วงจะมีการเปลี่ยนแปลงน้อยมากในแต่ละปีเนื่องจากในกลุ่มผู้ป่วยโรคความดันโลหิตสูงอาจจะไม่มีการเปลี่ยนแปลงน้ำหนักมากนักในช่วง 1-3 ปี แต่ทั้งนี้ทั้งนั้นในช่วง Obesity ก็ยังมีแนวโน้มที่จะสูงขึ้นตามลำดับดังในรูปที่ 4.11

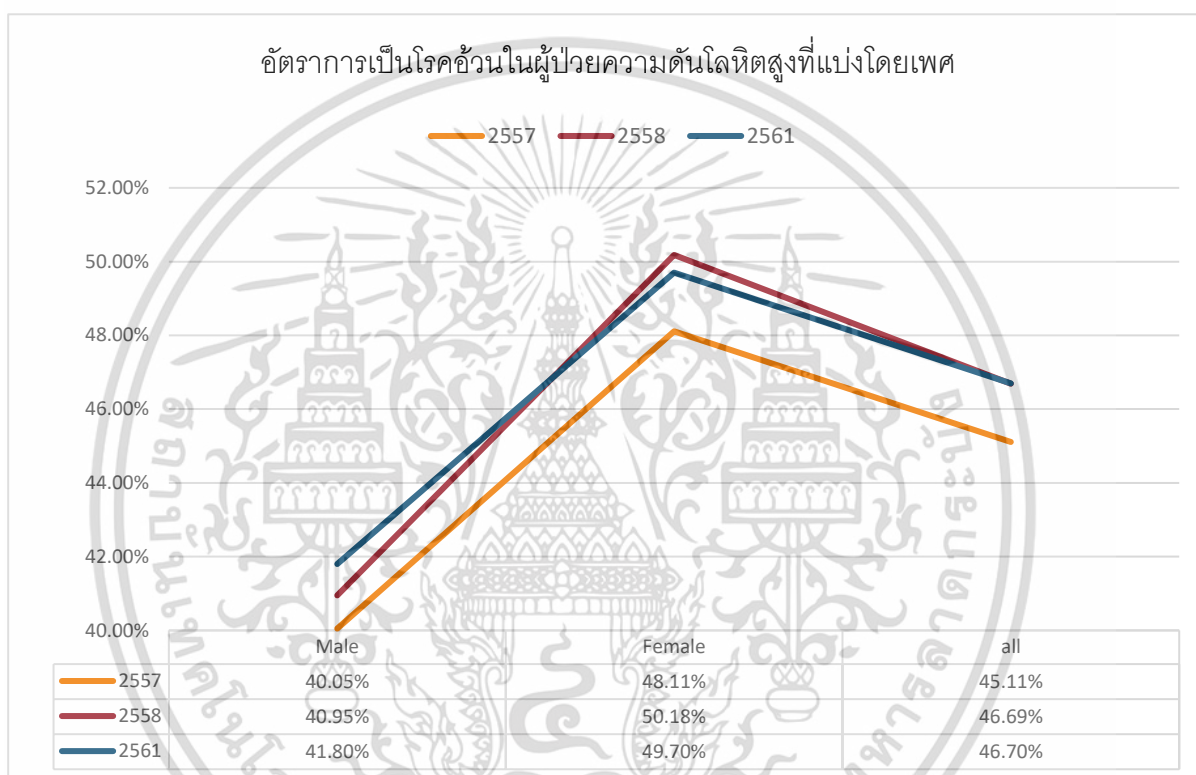


รูปที่ 4.11 กราฟแสดงอัตราการเปลี่ยนแปลงของ BMI Group

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4 อัตราการเป็นโรคอ้วนในผู้ป่วยความดันโลหิตสูงที่แบ่งโดยเพศในแต่ละปี

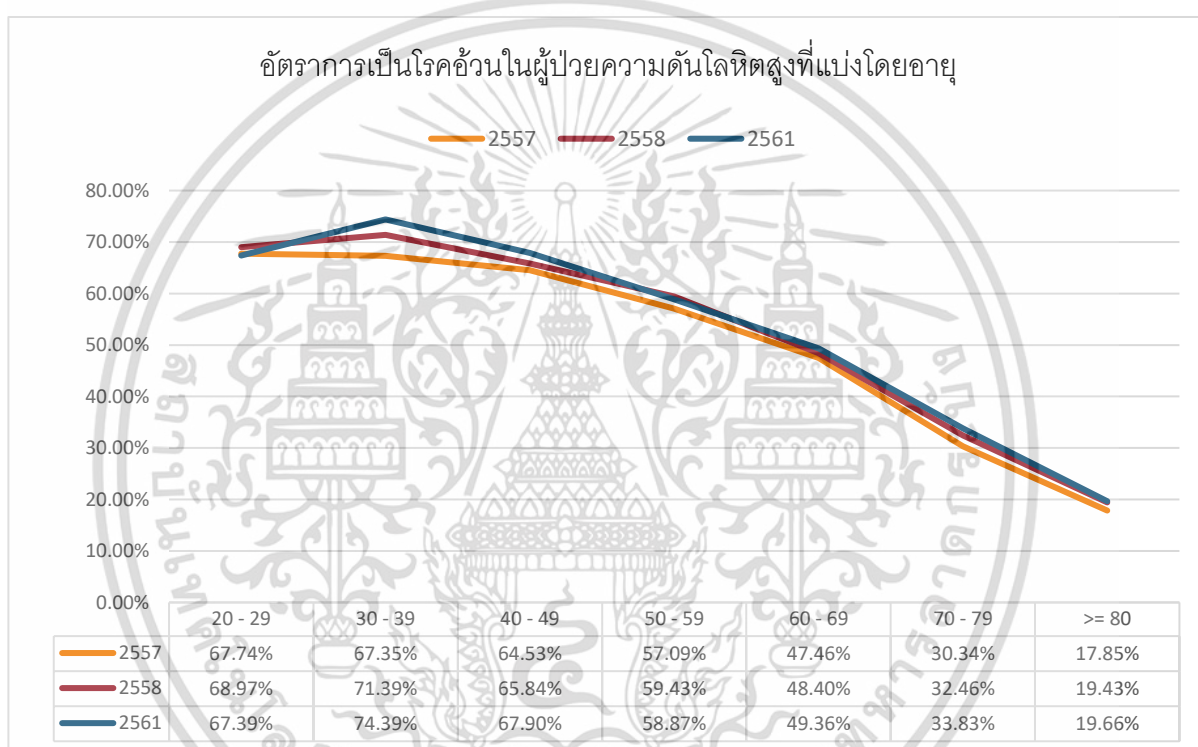
เมื่อนำจำนวนผู้ป่วยความดันโลหิตสูงที่เป็นโรคอ้วน ($BMI \geq 25 \text{ kg/m}^2$) มาเปรียบเทียบกับโดยเพศจะพบว่าทุกปีจะมีแนวโน้มที่มีผู้ป่วยเป็นโรคความดันโลหิตสูงและเป็นโรคอ้วนในเพศหญิงมากกว่าเพศชายร้อยละ 8.4 โดยประมาณ โดยโรคอ้วนของผู้ป่วยเบาหวานในเพศหญิงใกล้เคียงกันมากอยู่ที่ร้อยละ 50.18 และร้อยละ 49.70 ในปีพ.ศ. 2558 และพ.ศ. 2561 ตามลำดับ และในปีพ.ศ. 2557 มีอัตราการเป็นโรคอ้วนในผู้ป่วยความดันโลหิตสูงที่แบ่งโดยเพศต่ำที่สุดอยู่ที่เพศชายร้อยละ 40.05 เพศหญิงร้อยละ 48.11 เมื่อเทียบกับอีก 2 ปีแต่มีแนวโน้มไปในทิศทางเดียวกัน ดังรูปที่ 4.2



รูปที่ 4.12 กราฟแสดงอัตราการเป็นโรคอ้วนในผู้ป่วยความดันโลหิตสูงที่แบ่งโดยเพศ

4.5 อัตราการเป็นโรคอ้วนในผู้ป่วยความดันโลหิตสูงที่แบ่งโดยอายุในแต่ละปี

เมื่อนำจำนวนผู้ป่วยความดันโลหิตสูงที่เป็นโรคอ้วน ($BMI \geq 25 \text{ kg/m}^2$) มาเปรียบเทียบโดยอายุจะพบว่าช่วงอายุ 30-39 ปีมีจำนวนผู้ป่วยความดันโลหิตสูงและเป็นโรคอ้วนมากที่สุด โดยมีจำนวนผู้ป่วยความดันโลหิตสูงและเป็นโรคอ้วน 67.35%, 71.39% และ 74.39% ในปีพ.ศ. 2557 พ.ศ. 2558 และพ.ศ. 2561 ตามลำดับ และค่อยๆลดลงอย่างต่อเนื่องในทุกๆ 10 ปี บ่งชี้ให้เห็นว่าเมื่ออายุมากขึ้นค่า BMI ของผู้ป่วยความดันโลหิตสูงจะยิ่งลดลงดังในรูปที่ 4.13 สอดคล้องกับงานวิจัยที่กล่าวว่าเมื่ออายุมากขึ้นจะมีความเสี่ยงที่จะเป็นโรคอ้วนลดลง 1% ในทุกๆ 1 ปี (Adjust Odds Ratio = 0.99)

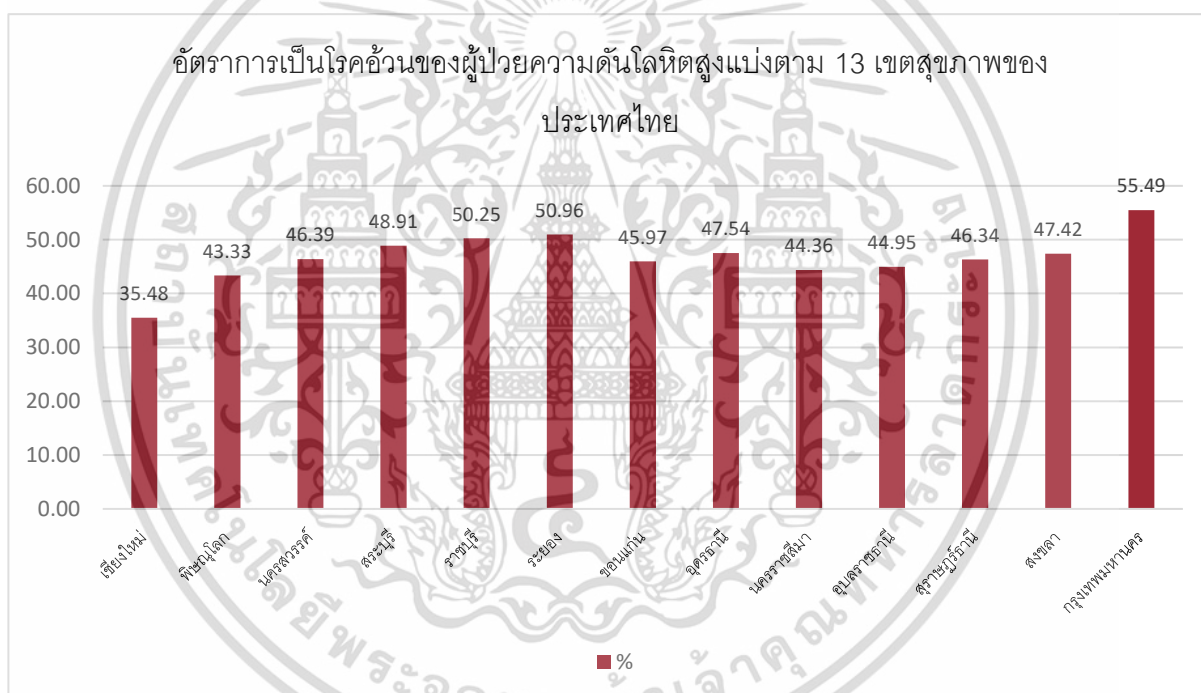


รูปที่ 4.13 กราฟแสดงอัตราการเป็นโรคอ้วนในผู้ป่วยความดันโลหิตสูงที่แบ่งโดยอายุ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6 อัตราการเป็นโรคอ้วนของผู้ป่วยความดันโลหิตสูงแบ่งตาม 13 เขตสุขภาพของประเทศไทย

ตามเขตสุขภาพ 13 เขตของประเทศไทยที่แบ่งตามความคล้ายคลึงของชีวิตความเป็นอยู่ของคนในกลุ่มจังหวัดเดียวกัน จะเห็นได้ว่าส่วนใหญ่กลุ่มผู้ป่วยความดันโลหิตสูงที่เป็นโรคอ้วนกระจายตัวอยู่ในทุกภูมิภาคของประเทศอย่างใกล้เคียงกัน แต่จะมีกลุ่มจังหวัดในภาคกลางคือเขตสุขภาพที่ 4 สระบุรี เขตสุขภาพที่ 5 ราชบุรี เขตสุขภาพที่ 6 ระยอง และเขตสุขภาพที่ 13 กรุงเทพมหานครซึ่งมีอัตราการเป็นโรคอ้วนของผู้ป่วยความดันโลหิตสูงมากกว่าเขตในภูมิภาคอื่นๆ สืบเนื่องมาจากความเจริญของเมือง อาหาร และการใช้ชีวิตของสังคมเมือง สอดคล้องกับงานวิจัยที่กล่าวว่าเมื่อผู้คนเข้าถึงอาหารจากอุตสาหกรรมแปรรูปได้มากขึ้นโดยอาหารเหล่านี้จะเป็นอาหารที่ให้พลังงานสูงเกินจำเป็นและนำไปสู่การเป็นโรคอ้วนได้ ดังในรูปที่ 4.14



รูปที่ 4.14 กราฟแสดงอัตราการเป็นโรคอ้วนของผู้ป่วยความดันโลหิตสูงแบ่งตาม 13 เขตสุขภาพ

4.7 การหาค่าทางสถิติเบื้องต้น

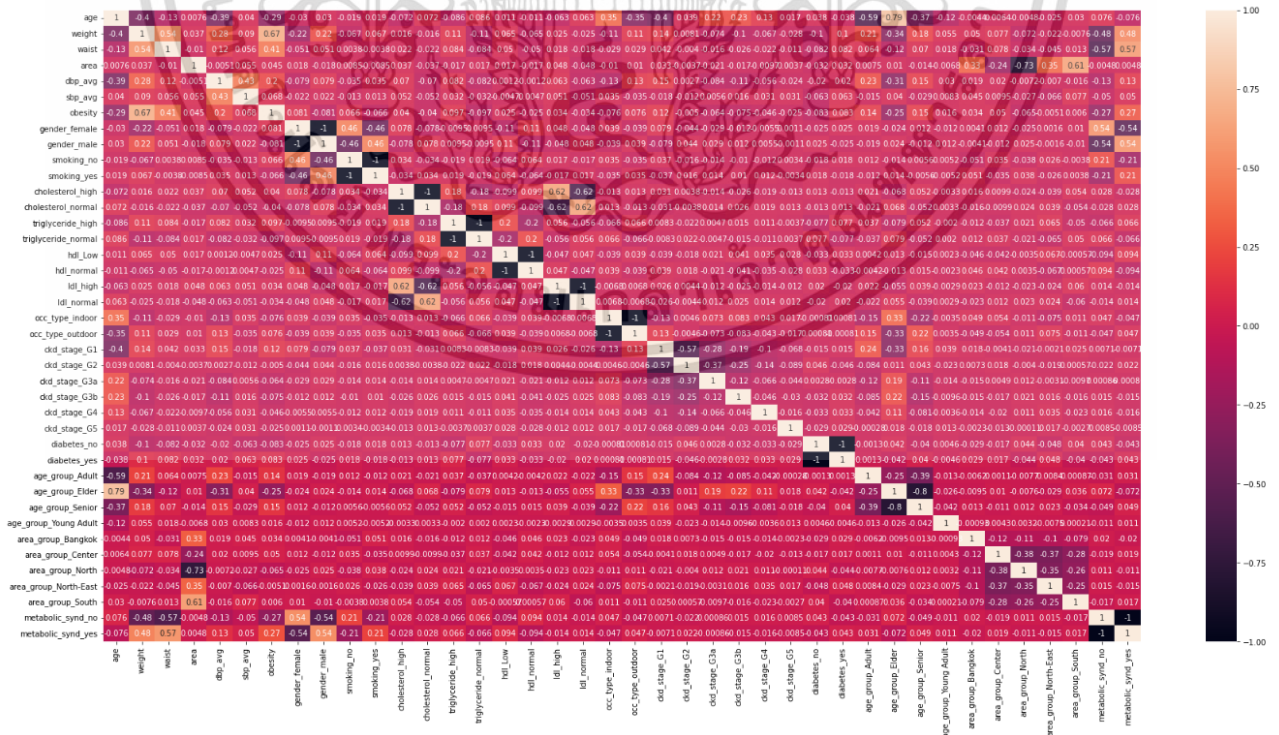
ค่าสถิติเบื้องต้นของตัวแปรที่เป็น Numerical รวมทั้ง 3 ปีจำนวน 79,663 records จะได้ดังตารางที่ 4.2

ตารางที่ 4.2 ค่าทางสถิติเบื้องต้นสำหรับข้อมูลที่เป็น Numerical

factors	mean	std	min	25	50	75	max
age	64.13	11.71	18.0	56.0	64.0	73.0	104.0
weight	62.22	13.08	23.0	53.0	61.0	70.0	160.0
waist	85.72	11.30	16.3	81.0	86.3	91.0	200.0
dbp_avg	76.25	8.40	40.0	70.3	76.7	81.7	130.5
sbp_avg	133.65	11.49	81.3	126.7	133.3	140.0	224.5

4.8 การหา correlation ระหว่างตัวแปรอิสระ

ขั้นตอนนี้จะเป็นการหาค่าตัวแปรที่สัมพันธ์กันเอง โดยจากตารางแสดง Heatmap ด้านล่างจะพบว่าค่า obesity สัมพันธ์กับค่า weight 0.67 เนื่องจากนำค่า Weight มาหา BMI ค่า gender_female สัมพันธ์กับ gender_male 0.54 ค่า cholesterol_high สัมพันธ์กับค่า ldl_high 0.62 และค่า cholesterol_normal สัมพันธ์กับค่า ldl_normal 0.62 จึงต้องตัดค่า weight และเลือกใช้เฉพาะ gender_male, cholesterol_high และ cholesterol_normal มาเข้าตัวแบบตามรูปภาพที่ 4.15



รูปที่ 4.15 กราฟ Heatmap แสดงความสัมพันธ์กันของตัวแปร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.9 การแสดงผลข้อมูลหลังจากนำเข้าตัวแบบ Logistic Regression

การแสดงผลข้อมูลหลังจากนำเข้าตัวแบบ Logistic Regression จะบ่งชี้ถึงแนวโน้มที่จะเกิดโรคอ้วนในแต่ละ Record ว่ามีความน่าจะเป็น (โรคอ้วน = 1 และไม่เป็นโรคอ้วน = 0) มากน้อยเพียงใด โดยมีตัวแปรอิสระทั้งหมดส่งที่มีความสัมพันธ์กับเหตุการณ์นั้นๆ ซึ่งเมื่อนำไปตรวจสอบประสิทธิภาพพบว่าตัวแปรอิสระทั้งหมด 13 กลุ่ม ที่ประกอบไปด้วย

- 1) age_group_Adult, age_group_Elder, age_group_Senior, age_group_Young Adult
- 2) area_group_Bangkok, area_group_Center, area_group_North, area_group_North-East, area_group_South
- 3) waist
- 4) sbp_avg
- 5) dbp_avg
- 6) gender_male
- 7) smoking_no, smoking_yes
- 8) cholesterol_high, cholesterol_normal
- 9) triglyceride_high, triglyceride_normal
- 10) hdl_Low, hdl_normal,
- 11) occ_type_indoor, occ_type_outdoor,
- 12) ckd_stage_G1, ckd_stage_G2, ckd_stage_G3a, ckd_stage_G3b, ckd_stage_G4, ckd_stage_G5
- 13) diabetes_no, diabetes_yes

จากนั้นนำค่าที่ได้มาคิดหา Accuracy, Precision และ Recall สำหรับการทำนายค่าการเป็นโรคอ้วน (obesity) ของทั้งตัวแบบได้ Accuracy เท่ากับ 74.68% Precision เท่ากับ 73.3% และ Recall เท่ากับ 71.6%

และแยกระหว่างเป็นและไม่เป็นโรคอ้วนได้ค่าดังนี้ค่า Precision ที่ Prediction = 0 (ไม่เป็นโรคอ้วน) อยู่ที่ 76% และ Prediction = 1 (เป็นโรคอ้วน) อยู่ที่ 73% ค่า Recall ที่ Prediction = 0 (ไม่เป็นโรคอ้วน) อยู่ที่ 78% และ Prediction = 1 (เป็นโรคอ้วน) อยู่ที่ 72% ค่า F1-score ที่ Prediction = 0 (ไม่เป็นโรคอ้วน) อยู่ที่ 77% และ Prediction = 1 (เป็นโรคอ้วน) อยู่ที่ 72% ดังในรูปที่ 4.16

	precision	recall	f1-score	support
0.0	0.76	0.78	0.77	12833
1.0	0.73	0.72	0.72	11066
accuracy			0.75	23899
macro avg	0.75	0.75	0.75	23899
weighted avg	0.75	0.75	0.75	23899

The Accuracy for Model is 74.6807976472276%

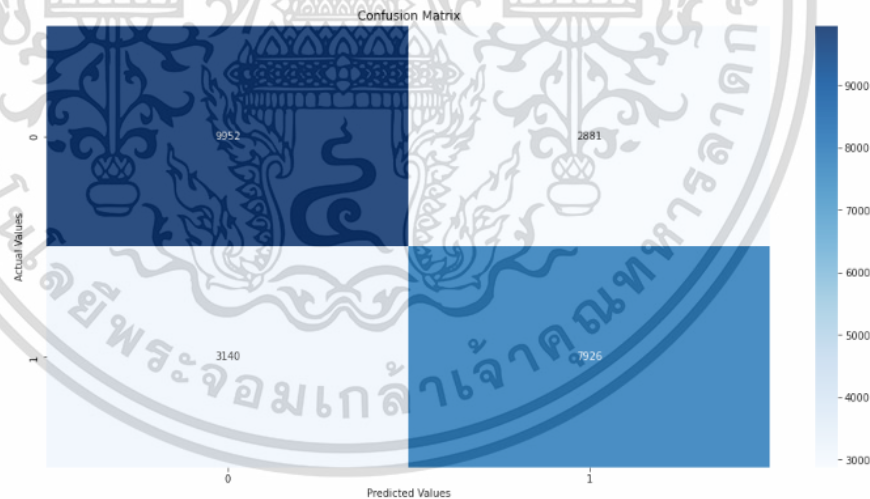
รูปที่ 4.16 ผลการแสดงค่า Accuracy, Precision, Recall และ f1-score ของตัวแบบ

4.10 ตาราง Confusion Matrix

ตาราง Confusion Matrix จะประกอบไปด้วย

- True Positive คือ ข้อมูลที่ตัวแบบทำนายว่าเป็นโรคอ้วนและตรงกับความจริง
- False Positive คือ ข้อมูลที่ตัวแบบต้องการทำนายว่าเป็นโรคอ้วนแต่ไม่เป็นโรคอ้วน
- False Negative คือ ข้อมูลที่ตัวแบบทำนายว่าเป็นไม่โรคอ้วนแต่เป็นโรคอ้วน
- True Negative คือ ข้อมูลที่ตัวแบบทำนายว่าไม่เป็นโรคอ้วนและตรงกับความจริง

จากผลที่ได้เราจะได้ข้อมูลที่ทำนายว่าผู้ป่วยไม่เป็นโรคอ้วนและตรงกับความจริงทั้งหมด 9,952 records และไม่ตรงกับความจริง (ผู้ป่วยเป็นโรคอ้วน) ทั้งหมด 3,140 records ในอีกกรณีจะได้ผู้ป่วยที่ทำนายว่าเป็นโรคอ้วนและตรงกับความจริงทั้งหมด 7,926 records และไม่ตรงกับความจริง (ผู้ป่วยไม่เป็นโรคอ้วน) ทั้งหมด 2,881 records ดังในรูปที่ 4.17



รูปที่ 4.17 ตาราง Confusion Matrix

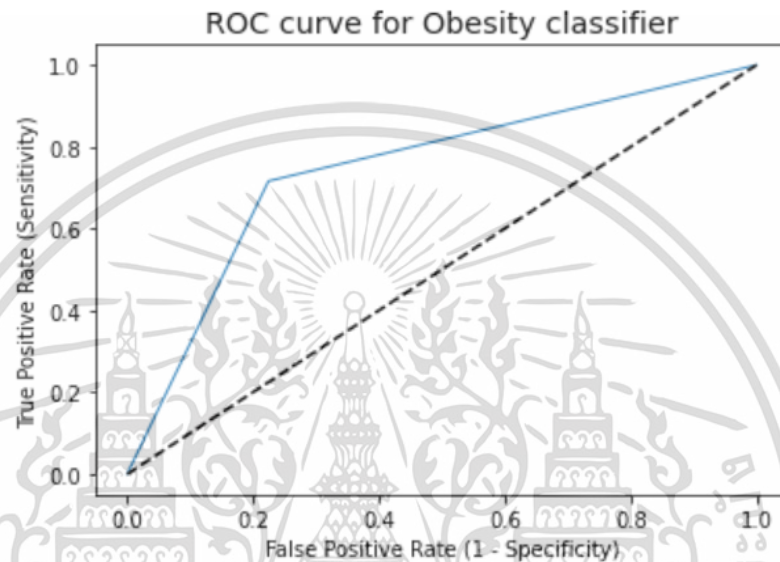
4.11 ค่า Mean Square Error และค่า Root Mean Square Error

ค่า Mean Square Error ที่ได้จะเท่ากับ 0.25 และค่า Root Mean Square Error ที่ได้จะเท่ากับ 0.50 ซึ่งแปลว่าตัวแบบทำนายค่าผิดไป 0.50 point

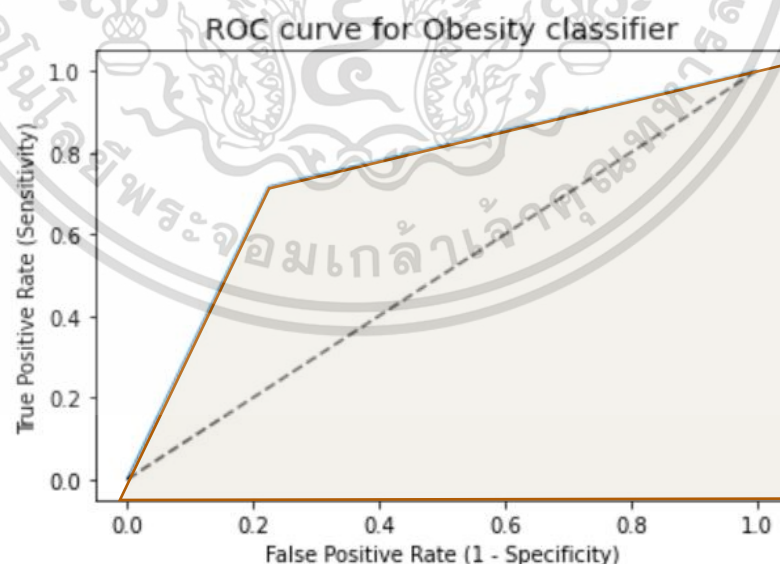
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.12 Receiver operating characteristic curve (ROC Curve)

เมื่อนำค่า True Positive Rate (Sensitivity) แกน y และ False Positive Rate (1-Specificity) แกน X มาสร้างกราฟจะได้กราฟ ROC Curve ออกมาดังในรูปที่ 4.18 และหาพื้นที่ใต้กราฟของกราฟ ROC Curve ได้ค่า 0.7459 ดังในรูป 4.19 จะแสดงให้เห็นถึงกลุ่ม y_{test} สามารถแบ่งกลุ่มของผู้ป่วยโรคความดันโลหิตสูงให้ผลถูกต้องตามความเป็นจริง 74.75%



รูปที่ 4.18 กราฟ ROC Curve



0.7458743145499255

รูปที่ 4.19 พื้นที่ใต้กราฟ ROC Curve

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.13 การแปลผลตัวแปรอิสระ (Multivariable factors) ที่สัมพันธ์ต่อตัวแปรตาม

จากการศึกษาในกลุ่มตัวอย่างผู้ป่วยความดันโลหิตสูงในประเทศไทยในปี พ.ศ. 2557 พ.ศ. 2558 และ พ.ศ. 2561 พบว่าแนวโน้มที่จะเกิดโรคอ้วนสูงขึ้นอย่างมีนัยสำคัญ (95% CI) โดยเพศเป็นหนึ่งในปัจจัยที่สำคัญต่อการเพิ่มขึ้นของอัตราการเป็นโรคอ้วนดังกล่าว ผู้ป่วยความดันโลหิตสูงในเพศหญิงมีความสัมพันธ์ที่จะเกิดโรคอ้วนมากกว่าในเพศชาย 5 เท่าอย่างมีนัยสำคัญเนื่องจากระบบเผาผลาญในร่างกายของเพศชายทำงานได้ดีกว่าเพศหญิงซึ่งจะสอดคล้องกับการศึกษาจากงานวิจัยที่เกี่ยวข้องบางส่วนในประเทศไทยและประเทศอื่นๆ เช่น อินเดีย เป็นต้น แต่ในบางประเทศก็ยังมีงานวิจัยที่แสดงผลว่าเพศชายมีความสัมพันธ์การเกิดโรคอ้วนมากกว่าในเพศหญิงสาเหตุเพราะปัจจัยร่วมจากนิสัยด้านอุปโภคบริโภค สังคมวัฒนธรรมและการใช้ชีวิตอื่นๆ ที่แตกต่างกันไปในแต่ละประเทศ

จากการระบุเพศที่มีความสัมพันธ์ต่อการเกิดโรคอ้วนในผู้ป่วยความดันโลหิตสูงแล้วเรายังพบว่าหากในเพศหญิงมีรอบเอวมากกว่า 90 cm หรือในเพศชายมีรอบเอวมากกว่า 80 cm ที่เรียกว่าโรคอ้วนลงพุง (Abdominal obesity) ซึ่งเป็นหนึ่งในโรคไม่ติดต่อเรื้อรัง จะทำให้มีความสัมพันธ์การเกิดโรคอ้วนเพิ่มสูงขึ้นเป็น 8.4 เท่า จากผู้ป่วยความดันโลหิตสูงที่รอบเอวปกติอย่างมีนัยสำคัญ

เป็นที่น่าสนใจเนื่องจากผลงานวิจัยอื่นๆ ที่เกี่ยวข้องเนื่องกล่าวว่าเมื่ออายุมากขึ้นความสัมพันธ์การเกิดโรคอ้วนจะน้อยลง 1% ต่อ 1 ปี แต่จากการศึกษาพบว่าในช่วงอายุ Young Adult (18-29 ปี) มีความเสี่ยงจะเป็นโรคอ้วนมากกว่าผู้ใหญ่ถึง 16% อย่างไม่มีนัยสำคัญอาจจะเพราะเนื่องจากข้อมูลในช่วง Young Adult นั้นมีจำนวนน้อยกว่าช่วงวัยอื่นๆ มาก แต่เมื่อกล่าวถึงช่วงวัย Senior และ Elder พบว่าความสัมพันธ์การเกิดโรคอ้วนลดลงเหลือ 56% และ 21% จากผู้ป่วยในช่วงอายุ Adult ตามลำดับ

ไขมันจากการบริโภคอาหารก็เป็นอีกหนึ่งปัจจัยที่มีความสัมพันธ์ต่อโรคอ้วนโดยตรง โดยที่คลอเรสเตอรอล (ไขมันไม่ดี) ซึ่งได้มาจากการรับประทานอาหารที่มีคลอเรสเตอรอลสูง อาทิเช่น อาหารทะเลจำพวกหมีกทะเล อาหารสำเร็จรูป เครื่องในสัตว์ เมื่อมีค่าสูงเกิดกำหนดจะทำให้มีความสัมพันธ์การเกิดโรคอ้วนเพิ่มขึ้น 8% และในทิศทางเดียวกันหากมี HDL (ไขมันดี) ซึ่งได้มาจากการบริโภคอาหารจำพวกปลาทะเลและไขมันจากพืชบางประเภทต่ำจะส่งผลให้ความสัมพันธ์การเกิดโรคอ้วนสูงขึ้นถึง 20% อย่างมีนัยสำคัญ

การแบ่งกลุ่มเขตสุขภาพออกเป็นภูมิภาคของประเทศไทยได้แก่ภาคเหนือ ภาคตะวันออกเฉียงเหนือ ภาคกลาง ภาคใต้ และกรุงเทพมหานครพบว่า ผู้ป่วยโรคความดันโลหิตสูงในกรุงเทพมหานครมีความสัมพันธ์การเกิดโรคอ้วนมากที่สุดเมื่อเทียบกับภาคอื่นๆ ตามมาด้วยภาคกลาง ภาคใต้ ภาคตะวันออกเฉียงเหนือ และภาคเหนือตามลำดับ อาจจะเป็นสาเหตุมาจากการใช้ชีวิตของชุมชนในเมือง ประเทศไทยกำลังพัฒนาไปสู่สังคมอุตสาหกรรมอย่างต่อเนื่อง การรับประทานอาหารจากอุตสาหกรรมแปรรูปที่มีพลังงานสูง และการออกกำลังกายที่ไม่เพียงพอสืบเนื่องต่อจากภาวะทางสังคมทำให้มีเวลาในการใส่ใจการรักษาสุขภาพน้อยลง เมื่อเทียบกับประชาชนที่อยู่ในภูมิภาคอื่นๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการศึกษาพบว่า การสูบบุหรี่ในผู้ป่วยความดันโลหิตสูงจะมีความสัมพันธ์ต่อการเกิดโรคอ้วนลดลง 13% อย่างมีนัยสำคัญเมื่อเทียบกับผู้ป่วยที่ไม่มีประวัติการสูบบุหรี่ ซึ่งจะสอดคล้องกับการศึกษาที่เกี่ยวข้องอื่นๆ ที่พบว่าค่า BMI ของผู้ที่สูบบุหรี่จะมีแนวโน้มต่ำกว่าผู้ที่ไม่สูบบุหรี่ เมื่อไปถึงศึกษาถึงข้อเท็จจริงพบว่าอาจจะเป็นผลกระทบของนิโคตินซึ่งไปกระตุ้นระบบการเผาผลาญในร่างกายของผู้ที่สูบบุหรี่ แต่ถึงแม้ว่าการสูบบุหรี่จะทำให้ลดอัตราการเกิดโรคอ้วนได้จริง แต่ก็ส่งผลกระทบต่อร่างกายในทางอื่น อาทิเช่นเกิดโรคมะเร็งปอด โรคถุงลมโป่งพอง รวมถึงปัญหาหลอดเลือดหัวใจทำให้โรคความดันโลหิตสูงแย่ลงได้

ปัจจัยในด้านอาชีพซึ่งถูกแบ่งออกเป็นการทำงานในร่ม เช่นการทำงานในออฟฟิศและทำงานกลางแจ้ง เช่นอาชีพรับจ้างต่างๆพบว่าอาชีพที่ทำงานในร่มมีความสัมพันธ์การเป็นโรคอ้วนลดลง 3% อย่างไม่มีนัยสำคัญ อาจจะเป็นสาเหตุมาจากกลุ่มคนที่ทำงานออฟฟิศมีโอกาสขยับร่างกายในแต่ละวันน้อยกว่าคนทำงานข้างนอกแต่ก็แทบจะไม่แตกต่างกันในด้านสถิติ

โรคเบาหวานเป็นผลมาจากการรับประทานอาหารเช่นกัน ในอาหารที่มีระดับน้ำตาลสูง อาทิเช่น เครื่องดื่มสำเร็จรูปชนิดต่างๆ ทำให้ระดับค่า FPG สูง ซึ่งผู้ป่วยโรคความดันโลหิตสูงและมีค่า FPG มากกว่าหรือเท่ากับ 126 มก.ดล. มากกว่า 2 ครั้งในการตรวจสอบจัดเป็นผู้ป่วยโรคเบาหวาน และเป็นปัจจัยทำให้เกิดโรคอ้วนได้เพิ่มสูงขึ้นถึง 45% อย่างมีนัยสำคัญเมื่อเทียบกับผู้ป่วยโรคความดันโลหิตสูงที่ไม่มีภาวะโรคเบาหวาน

โรคไตเรื้อรังเป็นโรคหนึ่งในบรรดาโรคแทรกซ้อนของโรคความดันโลหิตสูงอันเนื่องมาจากการทานอาหารรสจัดที่มีโซเดียมสูง พบว่าเมื่อลำดับความรุนแรงของโรคไตเรื้อรังสูงขึ้น (วัดจากอัตราการกรองของไตที่พบค่าโซเดียมสูงในปีสสว) จะทำให้ความสัมพันธ์กับการเป็นโรคอ้วนต่ำลงอย่างมีนัยสำคัญสืบเนื่องมาจากเมื่อโรคไตเรื้อรังรุนแรงขึ้นจะส่งผลให้ผู้ป่วยที่เป็นโรคความดันโลหิตสูงและโรคไตร่วมด้วยนั้นรับประทานอาหารไม่ค่อยได้ และเกิดอาการบวม น้ำโดยอาการดังกล่าวไม่ได้เป็นความอ้วนที่เกิดจากโรคอ้วนแต่เป็นผลมาจากการที่น้ำที่ควรอยู่ในเส้นเลือดออกมาอยู่ในเซลล์ร่างกายส่วนต่างๆแทนและไม่สามารถขับออกได้ ค่า Associated Multivariable Factors จะแสดงดังตารางที่ 4.3

ตารางที่ 4.3 ค่า Associated Multivariable Factors

Factors	BMI < 25 kg/m ²	BMI ≥ 25 kg/m ²	Adjusted Odds Ratio	95% CI	p-value
Age (years)					
Young Adult (18 - 29)	35 (31.8)	75 (68.2)	1.16	0.80 - 1.66	0.508
Adult (30 - 49)	2821 (33.1)	5702 (66.9)	1		
Senior (50 - 69)	20873 (47.1)	23432 (52.9)	0.56	0.53 - 0.58	<0.001
Elder (≥70)	19142 (71.6)	7583 (28.4)	0.21	0.20 - 0.22	<0.001
Gender					
Male	17709 (59.0)	12310 (41.0)	1		
Female	25162 (50.7)	24482 (49.3)	5.10	4.89 - 5.31	<0.001
Cholesterol (bad fats)					
Normal	28952 (55.2)	23456 (44.8)	1		
High	13919 (51.1)	13336 (48.9)	1.08	1.05 - 1.11	<0.001
HDL (good fats)					
Normal	34594 (54.4)	28949 (45.6)	1		
Low	8277 (51.3)	7843 (48.7)	1.20	1.16 - 1.24	<0.001
Smoking					
Smoking	9684 (60.3)	6364 (39.7)	1		
Never	33187 (52.2)	30428 (47.8)	1.13	1.09 - 1.17	<0.001
Occupation type					
Indoor	21063 (57.9)	15284 (42.1)	1		
Outdoor	21808 (50.3)	21508 (49.7)	0.97	0.95 - 1.00	0.125
Area					
Bangkok	1123 (44.5)	1400 (55.5)	1		
Center	11658 (50.0)	11673 (50.0)	0.68	0.63 - 0.74	<0.001
North	12374 (59.2)	8515 (40.8)	0.47	0.44 - 0.51	<0.001
North-East	10876 (54.2)	9172 (45.8)	0.57	0.52 - 0.61	<0.001
South	6840 (53.1)	6032 (46.9)	0.60	0.55 - 0.65	<0.001
Chronic Kidney Disease Stages					
G1	10836 (44.7)	13389 (55.3)	1		
G2	18562 (54.1)	15747 (45.9)	0.96	0.93 - 0.99	0.027
G3a	7464 (61.4)	4702 (38.6)	0.87	0.84 - 0.91	<0.001
G3b	4160 (66.7)	2075 (33.3)	0.78	0.74 - 0.83	<0.001
G4	1301 (68.5)	598 (31.5)	0.73	0.66 - 0.80	<0.001
G5	548 (66.1)	281 (33.9)	0.54	0.47 - 0.62	<0.001
Diabetes					
No	38698 (55.4)	31202 (44.6)	1		
Yes	4173 (42.7)	5590 (57.3)	1.45	1.39 - 1.51	<0.001
Abdominal Obesity					
No	27339 (67.0)	13466 (33.0)	1		
Yes	15532 (40.0)	23326 (60.0)	8.40	8.08 - 8.72	<0.001

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการดำเนินงานและข้อเสนอแนะ

5.1 สรุปผลการดำเนินงาน

ปัจจัยที่มีความสัมพันธ์กับโรคอ้วนในผู้ป่วยโรคความดันโลหิตสูงนั้นมีแนวโน้มทำให้ผู้ป่วยเป็นโรคอ้วนเพิ่มมากขึ้นอย่างต่อเนื่องแบบมีนัยสำคัญในทุกช่วงปีที่ศึกษาซึ่งหมายความว่าในอนาคตก็ยังคงมีแนวโน้มที่สูงขึ้น สืบเนื่องมาจากการใช้ชีวิตประจำวันของผู้ป่วยเป็นส่วนใหญ่ อาทิเช่น อาหาร การกิน การพักผ่อน การออกกำลังกาย ซึ่งทั้งหมดนี้นำไปสู่การเพิ่มขึ้นของน้ำหนักตัวและค่า BMI เมื่อลดปัจจัยเสี่ยงต่าง ๆ ลงได้ก็จะสามารถควบคุมโรคอ้วนในผู้ป่วยโรคความดันโลหิตสูงลงได้เช่นกัน นอกเหนือจากน้ำหนักตัวและค่า BMI แล้ว ยังมีปัจจัยอื่นๆที่สัมพันธ์กับการเกิดโรคอ้วนในผู้ป่วยความดันสูงได้แก่ เพศ อายุ ไขมันในเลือดทั้งชนิดดีและไม่ดี ประวัติการสูบบุหรี่ ภูมิภาคและโรคแทรกซ้อนอื่นๆ โดยเฉพาะ โรคเบาหวาน โรคไตวายเรื้อรัง และโรคอ้วนลงพุง

ถึงแม้ในปัจจุบันความก้าวหน้าทางสังคมยุคอุตสาหกรรมต่างๆที่พัฒนาอย่างไม่หยุดยั้งแต่กลับส่งผลในทางตรงกันข้ามกับสุขภาพโดยรวมผู้ป่วย ชีวิตในสังคมเมืองทำให้ผู้ป่วยโรคความดันโลหิตสูงนั้นมีปัจจัยในการเป็นโรคอ้วนเพิ่มสูงขึ้นและทำให้โรคความดันโลหิตสูงแย่ลง ดังนั้นจึงเป็นหน้าที่ของระบบสาธารณสุขท้องถิ่นที่ควรเข้าไปช่วยเหลือและส่งเสริมความรู้เรื่องการลดปัจจัยเสี่ยงในด้านต่างๆ โดยเฉพาะอย่างยิ่งในด้านการบริโภค การลดการบริโภคน้ำตาลทั้งในอาหารและเครื่องดื่มรวมถึงลดการรับประทานไขมันไม่ดีและเพิ่มการรับประทานไขมันดี รวมไปถึงการลดอาหารรสจัดที่มีโซเดียมสูง และจัดกิจกรรมส่งเสริมการออกกำลังกายเป็นระยะ โดยเริ่มจากในชุมชนซึ่งเป็นเส้นเลือดฝอยและขยายผลเป็นระดับภูมิภาคและระดับประเทศในที่สุด

5.2 ข้อจำกัดของการค้นคว้าอิสระ

หลังจากที่ทำการสรุปผลของการศึกษาแนวโน้มการเกิดโรคอ้วนในผู้ป่วยความดันโลหิตสูงพบข้อจำกัดดังนี้

- 1) ข้อมูลที่ได้มาเป็นข้อมูลจากโรงพยาบาลรัฐบาลเท่านั้น จึงไม่มีข้อมูลของผู้ป่วยโรคอ้วนที่รักษาที่โรงพยาบาลเอกชนหรือไม่ได้รับการรักษา
- 2) มีข้อจำกัดทางด้านข้อมูลเนื่องจากผู้ป่วยบางส่วนปฏิเสธการให้ข้อมูลหรือไม่มีข้อมูลถูกเก็บในเวชระเบียน (N/A) จึงต้องใช้วิธีการเก็บไว้เป็นค่าใดค่าหนึ่ง

5.3 ข้อเสนอแนะ

ข้อเสนอแนะของการศึกษาในครั้งนี้เป็นแนวทางในการนำเอาตัวแบบไปพัฒนาต่อซึ่งมีรายละเอียดดังนี้

1) เนื่องจากการเก็บข้อมูลของการศึกษาในครั้งนี้เป็นแบบ Stratified two-stage cluster sampling method with proportional to the size (PPS) ซึ่งจะสุ่มจากโรงพยาบาลและสุ่มผู้ป่วย ถ้าหากมีการสุ่มเพิ่มกลุ่มตัวอย่างให้กว้างมากขึ้นเช่นมีการสุ่มข้อมูลของผู้ป่วยเพิ่มขึ้นจากโรงพยาบาล เอกชน จะทำให้ผลลัพธ์ของตัวแบบมีความแม่นยำมากขึ้น

2) ในการศึกษาครั้งนี้เก็บตัวอย่างจากผู้ป่วยความดันโลหิตสูงช่วงปีพ.ศ. 2557 พ.ศ. 2558 และ พ.ศ. 2561 ซึ่งเป็นข้อมูลในอดีต จึงสามารถนำข้อมูลที่เป็นปัจจุบันมากกว่ามาใช้ในการทำนายได้ เพื่อให้ได้การพยากรณ์ที่แม่นยำและตรงกับสถานการณ์ของผู้ป่วยในปัจจุบันมากยิ่งขึ้น

3) หากนำไปศึกษาต่อในอนาคตสามารถนำผลการคาดการณ์ไปประยุกต์ใช้กับการทำโปรแกรม สุขภาพในการปรับเปลี่ยนการใช้ชีวิตเพื่อควบคุมปัจจัยต่างๆ เช่น ภาวะน้ำหนักเกิน น้ำตาลในเลือดสูง ไขมันในเลือดสูงให้ไม่แย่ลงไปกว่าเดิมหรือดีขึ้นสำหรับผู้ป่วยโรคความดันโลหิตสูงที่เป็นโรคอ้วนหรือหาวิธีป้องกันไม่ให้เกิดขึ้นสำหรับผู้ป่วยที่มีแนวโน้มกำลังจะเป็นโรคอ้วนในอนาคตร่วมกับทางสาธารณสุขในอนาคตได้

เอกสารอ้างอิง

1. กรกร ธรรมนำศีล, et al.. ผลของโปรแกรมการส่งเสริมการดูแลตนเอง ต่อพฤติกรรมของผู้สูงอายุโรคความดันโลหิตสูงที่ควบคุมไม่ได้ เขตกรุงเทพมหานคร. วารสารพยาบาลสาธารณสุข. พฤษภาคม – สิงหาคม 2558 ปีที่29 ฉบับที่2
2. Grundy SM, Hansen B, Smith SC, Jr., Cleeman JI, Kahn RA, American Heart A, et al. Clinical management of metabolic syndrome: report of the American Heart Association/National Heart, Lung, and Blood Institute/American Diabetes Association conference on scientific issues related to management. *Arterioscler Thromb Vasc Biol.* 2004;24(2):e19-24.
3. สถาบันวิจัยระบบสาธารณสุข. รายงานการสำรวจสถานะสุขภาพอนามัยของประชาชนไทย ด้วยการสอบถามและตรวจร่างกายทั่วประเทศ ครั้งที่ 4 [เอกสารอิเล็กทรอนิกส์]. [เข้าถึงเมื่อ 5 กรกฎาคม 2563]. เข้าถึงได้จาก <https://www.hiso.or.th/hiso5/report/report1.php>
4. Landsberg L, Aronne LJ, Beilin LJ, Burke V, Igel LI, Lloyd-Jones D, et al. Obesityrelated hypertension: pathogenesis, cardiovascular risk, and treatment: a position paper of The Obesity Society and the American Society of Hypertension. *J Clin Hypertens (Greenwich).* 2013;15(1):14-33.
5. จารุณี ปลายยอด, et al. ผลของโปรแกรมการส่งเสริมการมีส่วนร่วมในการดูแลตนเองต่อการรับรู้ ความสามารถในการดูแลตนเอง น้ำหนักตัว และการควบคุมความดันโลหิตของผู้ที่เป็นโรคความดันโลหิตสูง. *การพยาบาลอายุรศาสตร์-ศัลยศาสตร์.* Rama Nurs J • May - August 2012
6. Supinya Kongjarern. Obesity: Silent Killer in the Digital Era. Faculty of Nursing, Eastern Asia University. Vol. 11 No. 3 September-December 2017
7. Boonsub Sakboonyarat, et al. Trends, prevalence and associated factors of obesity among adults in a rural community in Thailand: serial crosssectional surveys, 2012 and 2018. *BMC Public Health (2020) 20:850*
8. ภาควิชาพยาธิวิทยา คณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร. วิธีการตรวจคัดกรองโรคเบาหวาน. [เอกสารอิเล็กทรอนิกส์]. [เข้าถึงเมื่อ 15 พฤษภาคม 2565]. เข้าถึงได้จาก http://www.med.nu.ac.th/pathology/patho_knowledge/file/แผนพับ%20การตรวจคัดกรองโรคเบาหวาน.pdf

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

9. National Kidney Foundation. Glomerular Filtration Rate (GFR). [เอกสารอิเล็กทรอนิกส์]. [เข้าถึงเมื่อ 20 พฤษภาคม 2565]. เข้าถึงได้จาก <https://www.hnapc.com/patient-education-links/glomerular-filtration-rate-gfr/>
10. HealthServ. 13 เขตสุขภาพของประเทศไทย. [เอกสารอิเล็กทรอนิกส์]. [เข้าถึงเมื่อ 5 พฤษภาคม 2565]. เข้าถึงได้จาก <https://healthserv.net/13เขตสุขภาพของประเทศไทย-1330>
11. Bumrungrad International Hospital. การจำแนกความดันโลหิตสูง. [เอกสารอิเล็กทรอนิกส์]. [เข้าถึงเมื่อ 25 พฤษภาคม 2565]. เข้าถึงได้จาก <https://www.bumrungrad.com/th/conditions/hypertension>
12. Jeong Uk Lim, Jae Ha Lee, J. S. Kim, Y. Hwang, Tae-Hyung Kim, S. Lim, K. Yoo, K. Jung, Young Kyoon Kim, C. Rhee, et al. Comparison of World Health Organization and Asia-Pacific body mass index classifications in COPD patients. Medicine. International Journal of Chronic Obstructive Pulmonary Disease. Published 21 August 2017
13. อำไพ ทองธีรภาพ. Unequal Probability Sampling. สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์
14. อรรวรรณ ศรีโสมพันธ์. โครงการฝึกอบรม "สร้างนักวิจัยรุ่นใหม่" (ลูกไก่). [เอกสารอิเล็กทรอนิกส์]. [เข้าถึงเมื่อ 21 พฤศจิกายน 2564]. เข้าถึงได้จาก https://www.ubu.ac.th/web/files_up/08f2015112820164117.pdf
15. Natdanai James. Supervised Learning (การเรียนรู้แบบมีผู้สอน) คืออะไร [เอกสารอิเล็กทรอนิกส์]. [เข้าถึงเมื่อ 18 พฤศจิกายน 2564]. เข้าถึงได้จาก <https://www.glugrgeek.com/education/ie311supervisedlearning/>
16. Yuth Kaiyawan. Principle and Using Logistic Regression Analysis for Research. วารสารวิจัยมหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย 4(1) : 1-12 (2555)
17. ญาณิศ อิงสะกาว, et al. Logistic Regression Models in Medical. สาขาสถิติ คณะวิทยาศาสตร์มหาวิทยาลัยขอนแก่น
18. Tim Bock. What are Dummy Variables? [เอกสารอิเล็กทรอนิกส์]. [เข้าถึงเมื่อ 15 พฤษภาคม 2565]. เข้าถึงได้จาก <https://www.displayr.com/what-are-dummy-variables/>
19. Pagon Gatchalee. Confusion Matrix เครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนาย ใน Machine learning [เอกสารอิเล็กทรอนิกส์]. [เข้าถึงเมื่อ 15 พฤษภาคม 2565]. เข้าถึงได้จาก <https://medium.com/@pagongatchalee/confusion-matrix-เครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนาย-ในMachine-learning-fba6e3f9508c>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

20. PCP. C&G DataCommunity. [เอกสารอิเล็กทรอนิกส์]. [เข้าถึงเมื่อ 15 พฤษภาคม 2565]. เข้าถึงได้จาก <https://medium.com/c-g-datacommunity/mse-rmse-mae-เลือกใช้อย่างไรดีมาดูที่ความหมาย-17b37b0b14b3>
21. Charlie Custer. 15 Python Libraries for Data Science You Should Know [เอกสารอิเล็กทรอนิกส์]. [เข้าถึงเมื่อ 18 พฤศจิกายน 2564]. เข้าถึงได้จาก <https://www.dataquest.io/blog/15-python-libraries-for-data-science/>
22. Suphakit Annopornchai. Python คืออะไร โปรแกรมภาษาไพธอน ใช้ทำอะไร [เอกสารอิเล็กทรอนิกส์]. [เข้าถึงเมื่อ 18 พฤศจิกายน 2564]. เข้าถึงได้จาก <https://saixiii.com/python-programming/>
23. Achieve.Plus. Rapidminer เสกคนไม่มีพื้นฐานให้เป็นเซียน [เอกสารอิเล็กทรอนิกส์]. [เข้าถึงเมื่อ 18 พฤศจิกายน 2564]. เข้าถึงได้จาก <https://medium.com/achieve-space/rapidminer-เสกคนไม่มีพื้นฐานให้เป็นเซียน-9bf6ab20d1aa>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ	นางสาวกุลยา วิสุทธิ์เทพ
วัน เดือน ปีเกิด	5 ตุลาคม พ.ศ. 2537
ที่อยู่ปัจจุบัน	9 เฉลิมพระเกียรติร.9 ซอย14 แยก12 หนองบอน ประเวศ กรุงเทพฯ 10250
ประวัติการศึกษา	2560 วิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ เกรดเฉลี่ย 2.32 สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้