รายงานการวิจัยฉบับสมบูรณ์

การประยุกต์ใช้โมเดลเครือข่ายแบบเบย์เพื่อวิเคราะห์ความเสี่ยงการเป็นโรคไม่ติดต่อจากการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย

Applying Bayesian Network for Chronic Diseases Risk Analysis from National Health Examination Survey in Thailand

ประเภททุน พัฒนานักวิจัยใหม่ รหัสโครงการ KREF015805

หัวหน้าโครงการ ผศ.ดร.กนกกรรณ์ ลี้โรจนาประภา

ได้รับการสนับสนุนเงินวิจัยจาก กองทุนวิจัย ประจำปี พ.ศ. 2559
คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รายงานการวิจัยฉบับสมบูรณ์

การประยุกต์ใช้โมเดลเครือข่ายแบบเบย์เพื่อวิเคราะห์ความเสี่ยงการเป็นโรคไม่ติดต่อจากการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย

Applying Bayesian Network for Chronic Diseases Risk Analysis from National Health Examination Survey in Thailand

ประเภททุน พัฒนานักวิจัยใหม่ รหัสโครงการ KREF015805

หัวหน้าโครงการ ผศ.ดร.กนกกรรณ์ ลี้โรจนาประภา

ได้รับการสนับสนุนเงินวิจัยจาก กองทุนวิจัย ประจำปี พ.ศ. 2559
คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อโครงการ การประยุกต์ใช้โมเดลเครือข่ายแบบเบย์เพื่อวิเคราะห์ความเสี่ยงการเป็นโรคไม่ติดต่อจากการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย
แหล่งเงิน กองทุนวิจัยสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ประจำปีงบประมาณ 2559 จำนวนเงินที่ได้รับการสนับสนุน 109,000 บาท
ระยะเวลาทำการวิจัย 1 ปี ตั้งแต่ 1 ตุลาคม 2558 ถึง 30 กันยายน 2559
ชื่อ-สกุล หัวหน้าโครงการ ผศ.ดร.กนกกรรณ์ ลี้โรจนาประภา ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

บทคัดย่อ

ในงานวิจัยนี้นำเสนอโมเดลเครือข่ายแบบเบย์ (Bayesian Network Model) เพื่อแสดงปัจจัยเชิงเหตุและผลต่อการเกิดโรคเบาหวานและโรคความดันโลหิตสูง ซึ่งเป็นโรคที่พบมากในกลุ่มโรคที่ไม่ติดต่อเรื้อรัง (Non-communicable Disease : NCD) โดยนำเสนอวัตถุประสงค์ ของงานวิจัยนี้ 2 ประการ คือ

1. วิเคราะห์ปัจจัยเสี่ยงหรือตัวบ่งชี้ความเสี่ยงที่สำคัญต่อการเกิดโรคไม่ติดต่อที่สำคัญของคนไทยจากโมเดลเครือข่ายแบบเบย์ (Bayesian Network Model) โดยวิธีจัดลำดับปัจจัยเสี่ยง และจัดกลุ่มความเสี่ยง

2. พยากรณ์การเกิดโรคไม่ติดต่อที่สำคัญของคนไทยด้วยโมเดลเครือข่ายแบบเบย์ (Bayesian Network Model)

ผู้วิจัยได้ทำการสร้างโมเดลเครือข่ายแบบเบย์ด้วยวิธีการสัมภาษณ์กับผู้เชี่ยวชาญหลายครั้งก่อนกำหนดโครงสร้างของโมเดลของโรคเบาหวานที่ประกอบด้วยตัวแปรทั้งสิ้น 8 ตัวแปร โครงสร้างของโมเดลของโรคความดันโลหิตสูงที่ประกอบด้วยตัวแปรทั้งสิ้น 10 ตัวแปร และทำการกำหนดค่าความน่าจะเป็นและความน่าจะเป็นแบบมีเงื่อนไขของแต่ละตัวแปรโดยอาศัยข้อมูลจากการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย ครั้งที่ 4 ก่อนทำการวิเคราะห์ตามแนวทางที่กำหนด

ผลจากการจัดลำดับความเสี่ยงโดยการเปรียบเทียบคำนวณค่า Normalised Likelihood และเปอร์เซ็นต์ที่เพิ่มขึ้นของแต่ละตัวแปร พบว่าปัจจัย 3 อันดับแรกสำหรับโรคเบาหวาน ประกอบด้วย ประวัติการเป็นโรคเบาหวานของครอบครัวสายตรง (Family History DM) การเป็นโรคอ้วน (Obesity) และปัจจัยด้านอายุ (Age) ในกลุ่มอายุ 35-59 ปี สำหรับปัจจัย 3 อันดับแรกสำหรับโรคความดันโลหิตสูง คือ ปัจจัยด้านภาวะโรคอ้วน (Obesity) ประวัติการเป็นโรคความดันโลหิตสูงของครอบครัวสายตรง (Family History HT) และปัจจัยด้านอายุ (Age) ในกลุ่มอายุ 35-59 ปี ซึ่งพบว่าเป็นกลุ่มตัวแปรเดียวกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การจัดกลุ่มความเสี่ยงช่วยให้ผู้วางแผนนโยบายสามารถกำหนดแนวทางในการวางนโยบายเพื่อจัดการกับความเสี่ยงของการเกิดโรคเบาหวานได้อย่างเหมาะสม ด้วยการวิเคราะห์ด้วยตารางประเมินความเสี่ยง (Risk Matrix) โดยแบ่งกลุ่มสาเหตุความเสี่ยงของการเกิดโรคจาก 2 องค์ประกอบหลักได้แก่ ความน่าจะเป็นของการเกิดความเสี่ยงของแต่ละปัจจัย และผลกระทบที่ในรูปของ % ของความน่าจะเป็นที่เพิ่มขึ้นของแต่ละปัจจัยเสี่ยงที่ 50%เมื่อทราบว่าเป็นผู้ป่วยเบาหวาน/ความดันโลหิตสูง (% Increase) แบ่งออกเป็นกลุ่ม 1 ความน่าจะเป็นและผลกระทบสูง กลุ่ม 2 ความน่าจะเป็นต่ำและผลกระทบสูง กลุ่ม 3 ความน่าจะเป็นสูงและผลกระทบต่ำ และกลุ่ม 4 ความน่าจะเป็นต่ำและผลกระทบต่ำ โดยพบว่าไม่มีปัจจัยใดอยู่ใน**กลุ่มที่ 1** ความน่าจะเป็นและผลกระทบสูง ส่วนในกลุ่มที่เหลือมีผลดังนี้ **กลุ่มที่ 2** ความน่าจะเป็นต่ำและผลกระทบสูง ในกลุ่มโรคเบาหวานได้แก่ ปัจจัยที่ด้านประวัติการเป็นโรคเบาหวานของครอบครัวสายตรง (Family History DM) และปัจจัยด้านการเป็นโรคอ้วน (Obesity) สำหรับโมเดลการเกิดโรคความดันโลหิตสูง ไม่มีโมเดลใดอยู่ในกลุ่มนี้ การจัดการลดความเสี่ยงให้กลับบุคคลกลุ่มนี้ทำได้โดยการ Transfer ความเสี่ยง เช่น แนะนำให้มีการทำประกันสุขภาพสุขภาพที่สามารถครอบคลุมการรักษาโรคเบาหวาน สำหรับ**กลุ่มที่ 3** ความน่าจะเป็นสูงและผลกระทบต่ำ ในกลุ่มโรคเบาหวานประกอบด้วย 2 ปัจจัยได้แก่ กลุ่มประชากรที่มีอายุ (Age) ในช่วง 35-59 ปี (ไม่รวมกลุ่มผู้สูงอายุ) พบว่าเป็นในกลุ่มประชากรส่วนใหญ่ (เนื่องจากสังคมไทยเข้าสู่สังคมสูงอายุ) และปัจจัยด้านการบริโภคผักและผลไม้ในปริมาณน้อย สำหรับกลุ่มโรคความดันโลหิตสูงประกอบด้วย 4 ปัจจัย ได้แก่ปัจจัยด้านอายุ (Age) มี ในกลุ่มอายุ 35-59 ปี ปัจจัยด้านการบริโภคเกลือและโซเดียม (Salt Consumption) ในระดับสูง ปัจจัยด้านที่อยู่อาศัย (Area of Residence) นอกเขตเทศบาล และสุดท้ายคือปัจจัยด้านเพศ (Sex) ในกลุ่มเพศหญิง การจัดการความเสี่ยงโดยเลือกวิธีการ Reduction/Mitigation สำหรับโรคเบาหวาน เช่น การสนับสนุนให้คนไทยบริโภคผักและผลไม้มากขึ้น และควรการจัดการความเสี่ยงกับโรคความดันโลหิตสูงจะเลือกวิธีการ Reduction/Mitigation เช่น การจัดการรณรงค์ให้ลดการบริโภคอาหารเค็มโดยเฉพาะในกลุ่มเพศหญิง อาศัยอยู่ในเขตเทศบาล ที่มีอายุ 35-59 ปี หรือให้มีการตรวจคัดกรองการเกิดโรคความดันโลหิตสูง สำหรับกลุ่มสุดท้าย**กลุ่มที่ 4** ความน่าจะเป็นต่ำและผลกระทบต่ำ ในกลุ่มโรคเบาหวาน ประกอบด้วย 3 ตัวแปรได้แก่ กลุ่มประชากรที่อาศัย (Area of Residence) อยู่ในเมือง หรือมีสถานะทางเศรษฐกิจและสังคม (Socio-Economic) ในระดับ 4 หรือมีกิจกรรมทางกาย (Physical Activity) ในระดับต่ำ สำหรับโมเดลโรคความดันโลหิตสูงประกอบด้วย 5 ตัวแปร ดังนี้ ปัจจัยด้านภาวะโรคอ้วน (Obesity) ปัจจัยด้านประวัติการเป็นโรคความดัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โลหิตสูงของครอบครัวสายตรง (Family History HT) ปัจจัยด้านการบริโภคผักและผลไม้ (Fruit And Vegetable Consumption) ในระดับปานกลางถึงมาก ปัจจัยด้านเคลื่อนไหวทางกาย (Physical Activity) ปัจจัยด้านสถานะทางสังคม (Socio-Economic) ในระดับ Wealth Quintile 4 การจัดการความเสี่ยงจะเลือกวิธีการ Retain/Absorption โดยผู้มีอำนาจตัดสินใจอาจไม่ต้องให้ความสำคัญกับปัจจัยเหล่านี้

ผลจากการพยากรณ์การเกิดโรคเบาหวานโดยพิจารณาจากค่า AUC พบว่า โมเดลมีความสามารถในการทำนายผลอย่างถูกต้องร้อยละ 76.70 และ 77.60 ซึ่งคำนวณจาก Training และ Testing Dataset ตามลำดับ ผลจากการพยากรณ์การเกิดโรคความดันโลหิตสูงมีค่า AUC เป็นร้อยละ 77.19 และ 73.71 ซึ่งคำนวณจาก Training และ Testing dataset ตามลำดับ

ในการเปรียบเทียบความสามารถของโมเดลโดยวิธีเครือข่ายแบบเบย์ อาจพิจารณาจากค่าต่างๆ ได้หลายค่า แต่งานวิจัยนี้เลือกใช้ค่าคะแนน F1 เป็นเกณฑ์ในการตัดสิน การใช้โมเดลเครือข่ายแบบเบย์สำหรับโรคเบาหวานจะให้ค่า F1 สูงสุดในกรณีที่กำหนดเกณฑ์ (Threshold) เป็น 0.1 มีค่าคะแนน F1 คือ 0.2449 โดยให้ค่าความถูกต้องของโมเดล (Accuracy) เป็น 80.50% เมื่อพิจารณาผลการพยากรณ์ของโมเดลเครือข่ายแบบเบย์ สำหรับโรคความดันโลหิตสูง พบว่ามีค่าคะแนน F1 สูงสุดเมื่อกำหนดระดับเกณฑ์ (Threshold) เป็น 0.2 ที่ทำให้มีค่า F1 เป็น 0.4412 โดยให้ค่าความถูกต้องของโมเดล (Accuracy) เป็น 65.65%

คำสำคัญ : โมเดลเครือข่ายแบบเบย์, โรคที่ไม่ติดต่อเรื้อรัง, การวิเคราะห์ความเสี่ยง, การพยากรณ์ความเสี่ยง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Research Title: Applying Bayesian Network for Chronic Diseases Risk Analysis from National Health Examination Survey in Thailand

Researcher: Kanogkan Leerojanaprapa

Faculty: Faculty of Science **Department:** Department of Statistics

ABSTRACT

Bayesian network model is proposed to capture causal factors for diabetes and hypertension. These diseases are very common in non-communicable Disease (NCD). The objectives of this research based on the Bayesian network model are as follows:

1. Prioritise risk factors or key risk indicators and classify risk factors for non-communicable disease of Thai people from the model.

2. Predict the prevalence of focused non-communicable disease of Thai People by using the model.

The Bayesian network model is constructed by interviewing experts several times before determining the structure of the diabetes model, which includes 8 variables. The structural model of the hypertensive model is composed of 10 variables. The configuration of the probability and the conditional probability of each variable is based on the 4th Thai National Health Examination Survey; before analysing the model according to the defined objectives.

The results from the risk prioritisation by comparing the Normalized Likelihood (NL) and percentage increase of each variable found that the top 3 factors for diabetes are Familial history (history of their family suffering from diabetes), Obesity and Age (in the 35-59 age group). The top three factors for hypertension are Obesity, Family History (history of their family suffering from hypertension), and Age in the 35-59 age group. We found that those variables belong to the same set.

The result of risk classification allows policy makers to set policy guidelines for managing the risks relevant to developing diabetes/hypertension. Analyses of the Risk

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Matrix, reveals two main components: the probability of each risk factor and the effect on the percentage increase of probability of each risk factor when known as positive of diabetes / high blood pressure. According to two criteria, the metric is divided into 4 groups by using 50% level as Group 1 which is high probability and high impact, Group 2 which is low probability and high impact, Group 3 which is high probability and low impact and Group 4 which is low probability and low impact. It is found that none of the factors are classified in **Group 1** (high probability and high impact). **Group 2** (low probability and high impact) for diabetes related to Familial history (history of their family suffering from diabetes) and obesity. For hypertensive models, it is found that none of those defined risk factors are in this group. Risk transfers, such as the introduction of health insurance that covers the treatment of diabetes are suggested to manage those risks. Next **Group 3** (high probability and low impact), for diabetes identified age (group 35-59 years), which is in the majority of Thai population, and low consumption of fruits and vegetables. For hypertension it consists of 4 factors: age (35-59 years), High salt consumption, Area of Residence, and Gender (female). Reduction / Mitigation is selected for diabetes, such as encouraging Thai people to consume more fruits and vegetables. Reduction/Mitigation might be done by campaigning to reduce salty food consumption especially for female living in urban area, aged 35-59 years and to have screening for hypertension. The last group, **Group 4** (low probability and low impact), for Diabetes consists of 3 factors: Residence in urban, Socio-Economic Level 4, and low level of physical activity. For Hypertension, it consists of 5 variables as obesity, history of hypertension in the family, fruit and vegetable consumption, physical activity, and Socio-Economic (the 4^{th} Quintile level). Retain/Absorption is applied for these risks so that the decision maker may not have to focus on them.

The results of the predictive model is summarised by the AUC and it shows that the model is able to predict 76.70% and 77.60% correctly for diabetes from training and testing dataset, respectively. For the predictive effect of high blood pressure the AUC values for hypertension model are 77.19%and 73.71% of correct prediction based on the training and testing dataset respectively.

On comparison of F1 score, the Bayesian network model for diabetes gives the highest F1 value at threshold of 0.1. The F1 score is 0.2449 (accuracy of 80.50%). When

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

considering the Bayesian network model for hypertension, the maximum F1 score is set at the threshold level of 0.2 and the F1 value is 0.4412 (accuracy of 65.65%).

Keywords : Bayesian Network Model, Non-communicable Disease (NCD), Risk Analysis, Risk Prediction

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับการสนับสนุนงบวิจัยจากทุนอุดหนุนการวิจัย จากทุนพัฒนานักวิจัยใหม่ กองทุนวิจัยสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ประจำปี 2559 นอกจากนี้ยังได้รับการสนับสนุนด้านข้อมูลจากสำนักงานการสำรวจสภาวะสุขภาพของประชาชนไทย (สสท.) ในการขออนุเคราะห์ข้อมูลจากฐานข้อมูลการสำรวจสุขภาพของประชาชนไทย โดยมีศาสตราจารย์นายแพทย์ วิชัย เอกพลากร หัวหน้าภาควิชาเวชศาสตร์ชุมชน คณะแพทยศาสตร์ โรงพยาบาลรามาธิบดี สนับสนุนและให้ความคิดเห็นด้านการสร้างโมเดลเครือข่ายแบบเบย์

ผศ.ดร.กนกกรรณ์ ลี้โรจนาประภา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

บทคัดย่อ ..... ก

ABSTRACT ..... ง

กิตติกรรมประกาศ ..... ช

สารบัญ ..... ซ

สารบัญตาราง ..... ฎ

สารบัญรูปภาพ ..... ฎ

บทที่ 1 บทนำ 1

1.1 ความสำคัญ และที่มาของปัญหา 1

1.2 วัตถุประสงค์ของโครงการ 3

1.3 ขอบเขตของการวิจัย 3

1.4 ประโยชน์ 4

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 5

2.1 ทฤษฎีเครือข่ายแบบเบย์ 5

2.1.1 การกำหนดโครงสร้างเครือข่ายแบบเบย์ 5

2.1.2 การกำหนดค่าความไม่แน่นอน – การกำหนดค่า โมเดลเครือข่ายแบบเบย์ 7

2.1.3 การอนุมานด้วยโมเดลเครือข่ายแบบเบย์ 8

2.1.4 การอนุมานโมเดลเครือข่ายแบบเบย์ที่มีความซับซ้อนด้วยความเป็นอิสระอย่างมีเงื่อนไข 9

2.1.5 Normalised likelihood (NL) 12

2.1.6 ความน่าจะเป็นภายหลังของแต่ละตัวแปร (Marginal Posterior Probability) 13

2.2 การวิเคราะห์ด้วยโมเดลเครือข่ายแบบเบย์ 14

2.3 การตรวจสอบความถูกต้องของตัวแบบในการพยากรณ์ 16

2.3.1 คอนฟูชั่น แมททริก (Confusion Matrix) 16

2.3.2 ค่า F1 18

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.3 เส้นโค้ง ROC และ AUC 19
2.4 วรรณกรรมที่เกี่ยวข้อง (Literature Review) 20
2.4.1 การใช้โมเดลเครือข่ายแบบเบย์กับข้อมูลความชุกการเกิดโรคในระดับประชากร 20
2.4.2 การใช้โมเดลเครือข่ายแบบเบย์ในการวิเคราะห์การเกิดโรค 21
บทที่ 3 วิธีการและแผนการดำเนินงานวิจัย 22
3.1 กรอบแนวความคิด 22
3.2 วิธีดำเนินการวิจัย 23
3.3 การวิเคราะห์จากโมเดลเครือข่ายแบบเบย์ 24
3.3.1 การอนุมานความน่าจะเป็นของการเกิดโอกาสที่ไม่พึงประสงค์ 25
3.3.2 จัดลำดับปัจจัยเสี่ยงที่อาจก่อให้เกิดโรคไม่ติดต่อที่สำคัญ 27
3.3.3 การจัดลำดับปัจจัยเสี่ยง (Risk Prioritisation) 28
3.3.4 การจัดกลุ่มปัจจัยเสี่ยง (Risk Classification) 28
3.3.5 การพยากรณ์การเกิดโรค 30
บทที่ 4 ผลการวิจัย 31
4.1 โรคเบาหวาน 31
4.1.1 ตัวแปรและนิยาม 31
4.1.2 โครงสร้างโมเดลเครือข่ายแบบเบย์ของการเกิดโรคเบาหวาน 33
4.1.3 การวิเคราะห์ปัจจัยเสี่ยงหรือตัวบ่งชี้ความเสี่ยงที่สำคัญต่อการเกิดโรคเบาหวาน 34
4.1.4 พยากรณ์การเกิดโรคเบาหวานของคนไทย 39
4.2 โรคความดันโลหิตสูง 44
4.2.1 ตัวแปรและนิยาม 44
4.2.2 โครงสร้างโมเดลเครือข่ายแบบเบย์ของการเกิดโรคความดันโลหิตสูง 46
4.2.3 การวิเคราะห์ปัจจัยเสี่ยงหรือตัวแปรบ่งชี้ความเสี่ยงที่สำคัญของการเกิดโรคความดันโลหิตสูงของคนไทย 47
4.2.4 พยากรณ์การเกิดโรคความดันโลหิตสูงของคนไทย 52

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5 สรุปผลการวิจัยและอภิปราย 57
5.1 ผลจากการวิเคราะห์ด้วยโมเดลเครือข่ายแบบเบย์ 57
5.1.1 การจัดลำดับความเสี่ยง 57
5.1.2 การจัดกลุ่มความเสี่ยง 58
5.1.3 การพยากรณ์ความเสี่ยง 60
5.2 เหตุผลที่วิธีโมเดลเครือข่ายแบบเบย์เหมาะสมสำหรับการศึกษาด้านปัจจัยเสี่ยงการเกิดโรคไม่ติดต่อเรื้อรัง 61
5.3 ข้อเสนอแนะ 62
บทที่ 6 สรุปผลผลิตที่ได้จากงานวิจัย 64
บรรณานุกรม 65
ภาคผนวก 68
ภาคผนวก ก. เอกสารหลักฐานอ้างอิงของผลผลิต 69
ภาคผนวก ข. โมเดลที่ได้จากการทำ Workshop 75
ภาคผนวก ค. สรุปการใช้จ่ายเงิน 81
ประวัตินักวิจัย 84

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตาราง 2-1 วิธีการวิเคราะห์ด้วยวิธีเครือข่ายแบบเบย์ ในวัตถุประสงค์ต่างๆ 15

ตาราง 2-2 คอนฟูชั่น แมททริก (Confusion Matrix) ในการจำแนก 2 กลุ่ม 16

ตาราง 2-3 สรุปงานวิจัยที่เกี่ยวข้องจำแนกตามจุดประสงค์ของการวิเคราะห์ด้วยโมเดลเครือข่ายแบบเบย์ 21

ตาราง 3-1 สรุปขั้นตอนการสร้างโมเดลเครือข่ายแบบเบย์จากข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย 23

ตาราง 4-1 ตัวแปร นิยาม และกลุ่มระดับตัวแปร (State) 32

ตาราง 4-2 สถานะของแต่ละตัวแปรที่ใช้ในการวิเคราะห์โรคเบาหวาน 34

ตาราง 4-3 ผลการเรียงลำดับปัจจัยเสี่ยงต่อการเกิดโรคเบาหวานของประชาชนไทย 36

ตาราง 4-4 คอนฟูชั่น แมททริก และค่าต่างๆ ที่ใช้ในการวัดความสามารถในการจำแนกโรคเบาหวาน ในระดับเกณฑ์ความน่าจะเป็นที่แตกต่างกัน 40

ตาราง 4-5 ค่า AUC จาก Training Dataset, Testing Dataset และ Full Dataset สำหรับโรคเบาหวาน 40

ตาราง 4-6 ตัวแปร นิยาม และกลุ่มระดับตัวแปร (State) 44

ตาราง 4-7 สถานะของแต่ละตัวแปรที่ใช้ในการวิเคราะห์ 47

ตาราง 4-8 ผลการเรียงลำดับปัจจัยเสี่ยงต่อการเกิดโรคความดันโลหิตสูงของประชาชนไทย 48

ตาราง 4-9 คอนฟูชั่น แมททริก และค่าต่างๆ ที่ใช้ในการวัดความสามารถในการจำแนกโรคความดันโลหิตสูง ในระดับเกณฑ์ความน่าจะเป็นที่แตกต่างกัน 53

ตาราง 4-10 ค่า AUC จาก Training และ Testing dataset และ Full Dataset สำหรับโรคความดันโลหิตสูง 53

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ

รูป 1-1 อัตราผู้ป่วยในต่อประชากรแสนคนจำแนกตามโรคไม่ติดต่อเรื้อรังที่สำคัญ ปี พ.ศ. 2546-2555 1

รูป 4-1 โครงสร้างโมเดลเครือข่ายแบบเบย์สำหรับการเกิดโรคเบาหวาน 33

รูป 4-2 ผลการจัดกลุ่มความเสี่ยงการเกิดโรคเบาหวาน 38

รูป 4-3 ROC Curve จาก Training Dataset สำหรับโรคเบาหวาน 41

รูป 4-4 ROC Curve จาก Testing Dataset สำหรับโรคเบาหวาน 42

รูป 4-5 ROC Curve จาก Full Dataset สำหรับโรคเบาหวาน 43

รูป 4-6 โครงสร้างโมเดลเครือข่ายแบบเบย์สำหรับการเกิดโรคความดันโลหิตสูง 46

รูป 4-7 ผลการจัดกลุ่มความเสี่ยงการเกิดโรคความดันโลหิตสูง 51

รูป 4-8 ROC Curve จาก Training Dataset สำหรับโรคความดันโลหิตสูง 54

รูป 4-9 ROC Curve จาก Testing Dataset สำหรับโรคความดันโลหิตสูง 55

รูป 4-10 ROC Curve จาก Full Dataset สำหรับโรคความดันโลหิตสูง 56

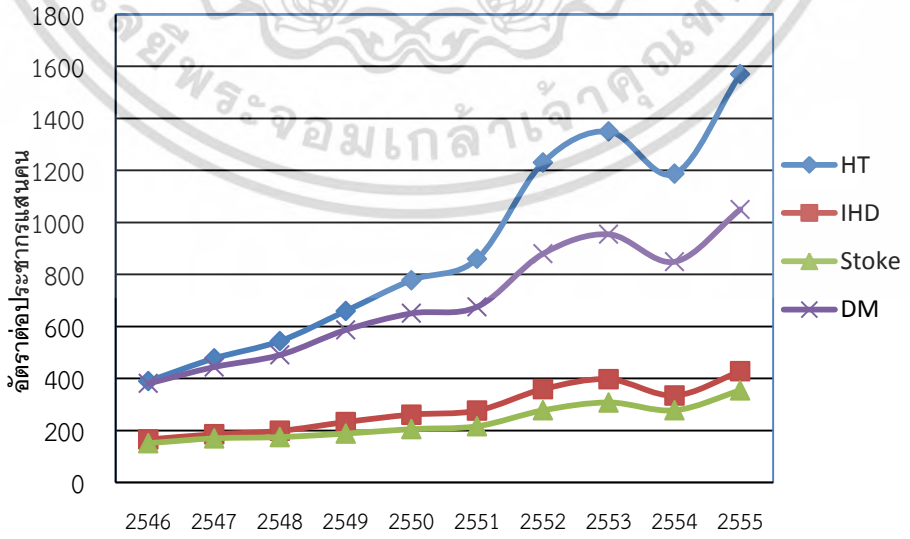
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1 บทนำ

ในบทนี้จะกล่าวถึงที่มาและความสำคัญในการทำการศึกษางานวิจัยนี้ ที่เกี่ยวกับการสร้างโมเดลเครือข่ายแบบเบย์ (Bayesian Network Model) สำหรับโรคไม่ติดต่อเรื้อรัง รวมทั้งระบุขอบเขตของงานวิจัย วัตถุประสงค์และประโยชน์ที่คาดว่าจะได้รับจากงานวิจัยนี้

1.1 ความสำคัญ และที่มาของปัญหา

โรคไม่ติดต่อเรื้อรัง (Chronic Disease) ที่สำคัญ ประกอบด้วย โรคหัวใจขาดเลือด (IHD) โรคความดันโลหิตสูง (HT) โรคเบาหวาน (DM) โรคหลอดเลือดสมองใหญ่ หรือ อัมพฤกษ์ อัมพาต (Stroke) โรคเหล่านี้มีแนวโน้มจำนวนผู้ป่วยสูงขึ้นและนำไปสู่อัตราการเสียชีวิตที่เพิ่มสูงขึ้น (อมรา ทองหงษ์, กมลชนก เทพสิทธา, & ภาคภูมิ จงพิริยะอนันต์, 2556) ดังรายงานการประจำปี 2556 ของสำนักงานโรคไม่ติดต่อ (2556) แสดงสถิติอัตราผู้ป่วยในด้วยโรคไม่ติดต่อที่สำคัญทั้ง 4 โรคดังกล่าว (ทั้งประเทศ) ปี พ.ศ. 2546–2555 นอกจากนี้ยังพบว่าสถานการณ์การป่วยในปี 2555 ต่อประชากร 100,000 คน ป่วยด้วยโรคความดันโลหิตสูง (HT) มากที่สุด คิดเป็น 1,570.6 ต่อประชากร 100,000 คน รองลงมาคือเบาหวาน (DM) คิดเป็น 1,050.1 ต่อประชากร 100,000 คน, โรคหัวใจขาดเลือด (IHD) คิดเป็น 427.5 ต่อประชากร 100,000 คนและ โรคหลอดเลือดสมองใหญ่ หรือ อัมพฤกษ์ อัมพาต (Stroke) คิดเป็น 354.5 ต่อประชากร 100,000 คน (ดังรูป 1-1)



ที่มา: รายงานประจำปี พ.ศ. 2556 (สำนักโรคไม่ติดต่อ กรมควบคุมโรค, 2556)

รูป 1-1 อัตราผู้ป่วยในต่อประชากรแสนคนจำแนกตามโรคไม่ติดต่อเรื้อรังที่สำคัญ ปี พ.ศ. 2546-2555

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กระทรวงสาธารณสุขจึงมุ่งเน้นนโยบายเพื่อลดอัตราการเกิดโรคดังกล่าวอย่างเป็นระบบเพื่อแก้ไขปัญหาทางด้านการแพทย์และสาธารณสุขอย่างยั่งยืนเพราะ ‘การป้องกันย่อมดีกว่าการรักษา: Prevention Is Better Than Cure’ เพราะการป้องกันจะช่วยในการลดต้นทุนการตรวจ การดูแลและการรักษาได้มากกว่า ถึงแม้ว่างานวิจัยทางการแพทย์ในเชิงลึกในรูปแบบการทดลองกับกลุ่มผู้ป่วยได้มีการศึกษาอย่างกว้างขวางทั้งในไทยและต่างประเทศ แต่การวิจัยเชิงลึกดังกล่าวอาจมุ่งเน้นด้านการรักษาเฉพาะกลุ่มผู้เป็นโรคนั้นๆ มากกว่าการป้องกัน สำหรับบางการศึกษาที่กำหนดขอบเขตการศึกษาเกี่ยวกับความชุกของโรคในกลุ่มประชาชนทั่วไปในบริบทของประชาชนไทยมักใช้ข้อมูลจากฐานข้อมูลคนไข้ที่เข้ามารับการรักษาในสถานพยาบาลต่างๆ แต่ข้อมูลดังกล่าวอาจไม่เป็นตัวแทนที่ดีของกลุ่มประชากรไทยทั้งหมดและยังคงขาดข้อมูลของผู้ป่วยที่ยังไม่ได้เข้าสู่ระบบของการรักษาหรือไม่ทราบว่าตนเป็นโรคนั้นๆ นอกจากนี้ฐานข้อมูลผู้ป่วยในโรงพยาบาลมักไม่มีข้อมูลที่เกี่ยวกับปัจจัยด้านพฤติกรรมความเสี่ยงของผู้ป่วยบันทึกไว้ ด้วยเหตุนี้ ‘การสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย’ โดยสำนักงานสำรวจสุขภาพประชาชนไทย สถาบันวิจัยระบบสาธารณสุข รวบรวมความชุกของโรค ปัจจัยเสี่ยงทางสุขภาพ และลักษณะทางประชากรที่สำคัญ ด้วยวิธีการสัมภาษณ์ ตรวจร่างกาย และการตรวจเลือดและปัสสาวะทางห้องปฏิบัติการเพื่อให้ได้ข้อมูลในทุกมิติ ข้อมูลจากการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกายเป็นฐานข้อมูลขนาดใหญ่ที่มีข้อมูลตัวบ่งชี้ทางชีวภาพ (Biomarker) ตัวบ่งชี้ด้านพฤติกรรม (Behavioural Marker) และตัวบ่งชี้ด้านลักษณะประชากร (Characteristic Marker) หากแต่การวิเคราะห์จากข้อมูลดังกล่าวมักเป็นการวิเคราะห์จากปัจจัยเสี่ยงเชิงเดี่ยวหรือจำแนกแต่ละปัจจัยเสี่ยงอย่างเป็นอิสระกันในรูปตารางจำแนกความถี่แบบ 2 หรือ 3 ทาง (ดูผลการวิเคราะห์จากรายงานการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย (สำนักงานสำรวจสุขภาพประชาชนไทย, 2552), (สำนักโรคไม่ติดต่อ กรมควบคุมโรค, 2556), (Aekplakorn Et Al., 2007)) ซึ่งอาจไม่สอดคล้องกับความซับซ้อนของการเกิดโรคไม่ติดต่อเรื้อรังที่มีสาเหตุที่ไม่ชัดเจนและมีสาเหตุเกิดจากหลายสาเหตุที่มีปฏิสัมพันธ์ (Interaction) ระหว่างปัจจัยสาเหตุต่างๆ ต่อการเกิดโรคดังกล่าว

จากรายงานการศึกษาและทบทวนวรรณกรรมพบว่า มีการวิเคราะห์ผลกระทบระหว่างปัจจัยเสี่ยงที่มีปฏิสัมพันธ์ (Interaction) เชิงปริมาณต่อภาวะของการเกิดโรคไม่ติดต่อต่างๆ เช่น ถ้าทราบว่ามีผู้ที่เป็นโรคอ้วนอาจส่งผลให้เกิดโรคเบาหวานและไขมันในเส้นเลือดสูงไปพร้อมๆ กัน หรือวิเคราะห์หาระดับปัจจัยเสี่ยงของคนที่เป็นโรคระบบหัวใจและหลอดเลือดอาจมีสาเหตุจากหลายปัจจัยได้แก่ ความดันโลหิตสูง เบาหวาน คลอเลสเตอรอลสูง โรคอ้วน และการสูบบุหรี่เป็นประจำ (สำนักงานสำรวจสุขภาพประชาชนไทย, 2552)

นอกจากนี้ยังพบว่าการทำงานร่วมกันระหว่างปัจจัยสามารถเป็นตัวเร่งให้เกิดโรคไม่ติดต่อต่างๆ มากขึ้น เช่น ถ้าทราบว่ามีผู้ที่เป็นโรคอ้วนและไม่ออกกำลังกาย โอกาสที่เขาจะเป็นโรคเบาหวานจะมากหรือน้อยกว่าผู้ที่เป็นโรคอ้วนและออกกำลังกาย ด้วยเหตุนี้การสร้างโมเดลในรูปความสัมพันธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ของเหตุและผล (Cause-Effect Relationship) ในรูปแบบเครือข่ายโดยโมเดลเครือข่ายแบบเบย์ (Bayesian Network Model) สามารถนำมาประยุกต์ใช้กับกับปัญหาความสัมพันธ์ระหว่างปัจจัยเสี่ยงต่างๆ ต่อการเกิดโรคไม่ติดต่อที่สำคัญ และยังสามารถแสดงผลการวิเคราะห์ในเชิงลึกเพื่อจัดเรียงลำดับความสำคัญของปัจจัยเสี่ยงเพื่อเพิ่มประสิทธิภาพการวิเคราะห์สาเหตุของโรคไม่ติดต่อและนำไปสู่การวางแผนจัดการปัจจัยเสี่ยงเหล่านั้นอย่างมีประสิทธิภาพต่อไป

ผู้วิจัยคาดหวังว่าผลการวิเคราะห์จากโมเดลจะช่วยวิเคราะห์หาสาเหตุหลักของการเกิดโรคไม่ติดต่อดังกล่าวอย่างเป็นระบบในบริบทของประเทศไทยซึ่งจะสามารถตอบสนองกรอบการวิจัยทางการแพทย์และสาธารณสุขในการวางแผนการป้องปรามโรคไม่ติดต่อที่เหมาะสมกับสภาวะแวดล้อมในปัจจุบันของประทศไทย โดยมุ่งเน้นการแก้ไขที่ต้นเหตุของโรคที่แท้จริงซึ่งจะเป็นการจัดการกับปัจจัยเสี่ยงของโรคได้อย่างมีประสิทธิภาพอย่างแท้จริง

1.2 วัตถุประสงค์ของโครงการ

การวิจัยในครั้งนี้มีจุดประสงค์ดังต่อไปนี้

1. วิเคราะห์ปัจจัยเสี่ยงหรือตัวบ่งชี้ความเสี่ยงที่สำคัญต่อการเกิดโรคไม่ติดต่อที่สำคัญของคนไทยจากข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกายด้วยโมเดลเครือข่ายแบบเบย์ (Bayesian Network Model)
2. พยากรณ์การเกิดโรคไม่ติดต่อที่สำคัญของคนไทยจากข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกายด้วยโมเดลเครือข่ายแบบเบย์ (Bayesian Network Model)

1.3 ขอบเขตของการวิจัย

โรคไม่ติดต่อ ที่นำมาศึกษาครั้งนี้ประกอบด้วย โรคเบาหวาน (Diabetes) และโรคความดันโลหิตสูง (Hypertension) โดยพิจารณาความเหมาะสมของการสร้างโมเดลจากข้อมูลการสำรวจสุขภาพของประชาชนไทยโดยการตรวจร่างกาย

กลุ่มประชากรไทย อายุ 15 – 59 ปี เท่านั้น

ข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย มีการดำเนินการสำรวจโดยการสุ่มตัวอย่างคนไทยในช่วงอายุต่างๆ เพื่อศึกษาความชุกของปัญหาสุขภาพของประชาชนไทยโดยทำการเก็บข้อมูลด้วยวิธีการสัมภาษณ์และตรวจร่างกายจากกลุ่มตัวอย่างในช่วงอายุต่างๆ ที่อาศัยอยู่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทุกภาคของประเทศไทย การสำรวจเริ่มต้นครั้งแรกในปี พ.ศ. 2534 และดำเนินการสำรวจอย่างต่อเนื่องทุก 5 ปี การสำรวจครั้งล่าสุดคือครั้งที่ 4 ปี พ.ศ. 2551-2552[1]

1.4 ประโยชน์

การศึกษาคาดว่าจะสามารถสร้างประโยชน์โดยช่วยให้ผู้บริหารด้านนโยบายสาธารณสุขสามารถวิเคราะห์หาสาเหตุหลักของปัญหาการเกิดโรคไม่ติดต่อจนนำไปสู่การกำหนดนโยบายจัดการปัจจัยเสี่ยงเพื่อเตรียมแผนป้องกัน ควบคุม ส่งเสริมสุขภาพและการรักษาอย่างมีประสิทธิภาพ

[1] ครั้งล่าสุดเป็นครั้งที่ 5 แต่ข้อมูลกำลังอยู่ระหว่างประมวลผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

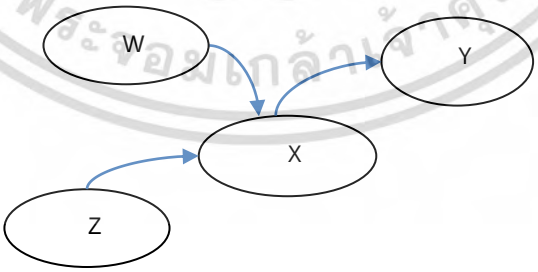
ในบทนี้จะนำเสนอแนวทางการประยุกต์ใช้ทฤษฎีเครือข่ายเบย์ (Bayesian Network Model) มาอธิบายการเกิดโรคไม่ติดต่อเรื้อรัง เพื่อนำไปสู่ผลจากการวิเคราะห์ความเสี่ยงของการเกิดโรคไม่ติดต่อเรื้อรังที่จะกล่าวต่อไปในบทที่ 4 การนำเสนอในบทนี้จะเริ่มต้นด้วยการกล่าวถึงทฤษฎีโมเดลเครือข่ายแบบเบย์ที่เกี่ยวข้องพร้อมตัวอย่างประกอบในหัวข้อ 2.1 จากนั้นจึงได้นำเสนอวิธีการตรวจสอบและเปรียบเทียบโมเดลว่ามีความสามารถในการพยากรณ์มากน้อยเพียงใดโดยใช้หลักเกณฑ์ต่างๆ ในการเปรียบเทียบผลจากการพยากรณ์ที่ได้จากโมเดลและค่าที่แท้จริงในหัวข้อ 2.2 และในหัวข้อสุดท้ายเป็นการนำเสนอการทบทวนวรรณกรรมในกรณีที่มีการนำโมเดลเครือข่ายแบบเบย์มาใช้วิเคราะห์ที่เกี่ยวข้องกับงานวิจัยนี้

2.1 ทฤษฎีเครือข่ายแบบเบย์

2.1.1 การกำหนดโครงสร้างเครือข่ายแบบเบย์

โมเดลเครือข่ายแบบเบย์คือ โมเดลที่แสดงความสัมพันธ์ระหว่างตัวแปรหรือปัจจัยที่มีความซับซ้อน โดยรูปแบบความสัมพันธ์สามารถกำหนดได้หลายรูปแบบ ขึ้นอยู่กับผู้วิจัยจะใช้หลักเกณฑ์ใดในการกำหนดโครงสร้างของโมเดล

ตัวอย่าง 2-1 ตัวอย่างโมเดลเครือข่ายแบบเบย์



รูป 2-1 ตัวอย่างโมเดลเครือข่ายแบบเบย์ ที่ประกอบด้วย 4 ตัวแปร

โดยทั่วไปแล้วโมเดลเครือข่ายแบบเบย์จะมีความซับซ้อนมากขึ้นเมื่อเป็นโครงสร้างของตัวแปรหลายตัวและทำให้มีลูกศรจำนวนมากเชื่อมโยงในโมเดล โดยในงานวิจัยนี้จะขออธิบายการกำหนดตัว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

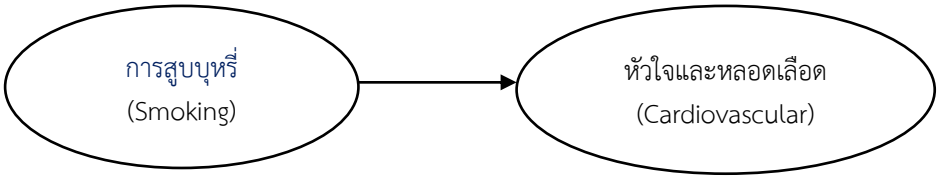
แปรที่นำมาใช้ในโครงสร้างโมเดลเครือข่ายแบบเบย์ในรูปความสัมพันธ์เชิงเหตุและผล สามารถแบ่งกลุ่มตัวแปรเป็น 3 กลุ่ม ดังนี้คือ

1. ตัวแปรที่เป็นต้นเหตุของปัญหา (Root Cause) คือ ตัวแปรที่ไม่มีตัวแปรอื่นอยู่ก่อนหน้าหรือไม่มีลิงค์จากตัวแปรอื่นพุ่งเข้าตัวแปรต้นเหตุ จากตัวอย่างข้างต้นตัวแปร **W** และ **Z** เป็นตัวแปรต้นเหตุของปัญหา (Root Cause)
2. ตัวแปรที่เป็นสาเหตุ (Cause Variable) คือตัวแปรที่มีลิงค์จากตัวมันเองพุ่งไปยังตัวแปรอื่น ซึ่งอาจตัวแปรต้นเหตุ (Root Cause) หรือไม่ก็ได้ บางครั้งอาจเรียก Parent Variable หากมีความสัมพันธ์ในรูปแบบใดๆ ที่ไม่จำเป็นต้องมีความสัมพันธ์เชิงเหตุและผล จากตัวอย่างข้างต้น ตัวแปรที่เป็นสาเหตุที่เป็นสาเหตุ ประกอบด้วย **W** และ **Z** เป็นสาเหตุของ **X** และ ตัวแปร **X** เป็นสาเหตุของตัวแปร **Y**
3. ตัวแปรผลกระทบ (Effect Variable) คือตัวแปรที่มีลิงค์จากตัวแปรอื่นพุ่งเข้า บางครั้งอาจเรียก Child Variable หากมีความสัมพันธ์ในรูปแบบใดๆ ที่ไม่จำเป็นต้องมีความสัมพันธ์เชิงเหตุและผล และหากเป็นตัวแปรสุดท้ายที่ไม่มีลิงค์พุ่งออกจากตัวแปรนั้น ในงานวิจัยนี้จะเรียกว่า ตัวแปรเป้าหมาย (Focal Variable) จากตัวอย่างข้างต้น ตัวแปร Y ตัวแปรเป้าหมาย (Focal Variable) ของโมเดลที่ใช้ในงานวิจัยนี้

โมเดลที่สามารถนำเสนอในรูปแบบความสัมพันธ์ระหว่างเหตุและผล ซึ่งมีลักษณะเช่นเดียวกับสาเหตุการเกิดโรคต่างๆ ได้ ในรูปแบบของกราฟที่ใช้ลูกศรแสดงความสัมพันธ์จากปัจจัยที่เป็นต้นเหตุสู่ปัจจัยที่เป็นผลลัพธ์ที่ทำให้เกิดโรคต่างๆ

การกำหนดโครงสร้างความสัมพันธ์ของการเกิดโรคด้วยโมเดลเครือข่ายแบบเบย์ แสดงด้วยตัวอย่าง อย่างง่ายดังตัวอย่าง 2-2

ตัวอย่าง 2-2 การสูบบุหรี่เป็นสาเหตุของการเกิดโรคหัวใจและหลอดเลือด (Cardiovascular) สามารถแสดงได้ด้วยกราฟ ดังรูป 2-2



รูป 2-2 ตัวอย่างโครงสร้างโมเดลเครือข่ายแบบเบย์ (อย่างง่าย)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

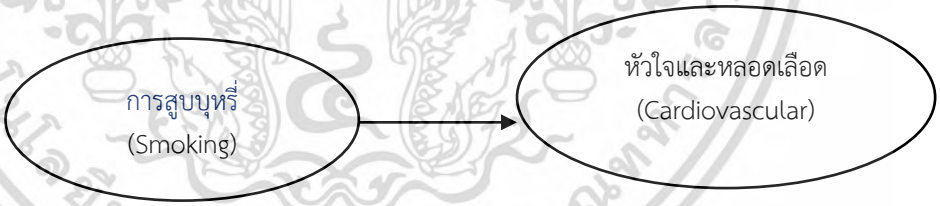
2.1.2 การกำหนดค่าความไม่แน่นอน – การกำหนดค่า โมเดลเครือข่ายแบบเบย์

ความน่าจะเป็น (Probability) ซึ่งคือ โอกาสของแต่ละสถานะ (State) ของตัวแปร ซึ่งบ่งบอกถึงระดับของความเป็นไปได้ที่มีค่าระหว่าง 0 ถึง 1 เพื่อแสดงความน่าจะเป็นของการเกิดขึ้นในแต่ละสถานะ (State) สำหรับแต่ละตัวแปร

ตัวอย่าง 2-3 "การสูบบุหรี่ (Smoking)" เป็นตัวแปรต้นเหตุของปัญหา (Root Cause) ของ “การเกิดโรคหัวใจและหลอดเลือด (Cardiovascular)” ที่เป็นตัวแปรผลลัพธ์ (Output) โดยที่ทราบว่ามีเพียงบางส่วนของประชาการที่สูบบุหรี่ ซึ่งสมมุติว่ามีร้อยละ 19.9 ของประชากร แทนด้วย $P(X=x)=P(X=smoking)$ แสดงดังตารางซ้ายมือในรูป 2-3

นอกจากนี้ยังสามารถแสดงผลกระทบที่ไม่แน่นอนของความสัมพันธ์ระหว่างปัจจัยได้ด้วย เช่น ไม่จำเป็นที่คนสูบบุหรี่ทุกคนจะเป็นโรคหัวใจและหลอดเลือด ด้วยเหตุนี้ระดับความสัมพันธ์ระหว่างปัจจัยทั้งสองสามารถกำหนดได้ด้วยระดับความน่าจะเป็นหรือสัดส่วนของผู้ที่สูบบุหรี่และเป็นโรคหัวใจและหลอดเลือดต่อจำนวนผู้สูบบุหรี่ทั้งหมด จากรูปแบบผลลัพธ์ที่ไม่แน่นนอน (Uncertain Effect) สามารถแสดงได้ด้วยตารางความน่าจะเป็นแบบมีเงื่อนไข (Conditional Probability) ซึ่งเป็นตารางทางขวามือ แสดงดังรูป 2-2 เช่นพบว่า ร้อยละ 34.1 ของผู้สูบบุหรี่เป็นโรคหัวใจและหลอดเลือดแทนด้วย

$$P(Y=y|X=x)=P(Y=\text{positive}\,|\,X=\text{smoking})=0.341$$



การสูบบุหรี่ (Smoking)	ความน่าจะเป็น
สูบ (Smoking)	0.199
ไม่สูบ (Non-smoking)	0.801

หัวใจและหลอดเลือด (Cardiovascular)	การสูบบุหรี่ (Smoking)	
	สูบ	ไม่สูบ
เป็นโรค (Positive)	0.341	0.079
ไม่เป็นโรค (Negative)	0.659	0.921

หมายเหตุ : ผลรวมแนวคอลัมน์เป็น 1

รูป 2-3 ตัวอย่างการกำหนดค่าความน่าจะเป็นของโมเดลเครือข่ายแบบเบย์ ด้วยตารางความน่าจะเป็น (Probability Table : PT) และตารางความน่าจะเป็นแบบมีเงื่อนไข (Conditional Probability Table : CPT)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยทั่วไปแล้วตัวแปรสามารถอธิบายได้จากตัวแปรต้นเหตุของปัญหา (Root Cause) และตัวแปรผลลัพธ์ (Effect) ซึ่งสามารถวัดได้จากตารางความน่าจะเป็น (Probability Table : PT) และตารางความน่าจะเป็นแบบมีเงื่อนไข (Conditional Probability Table : CPT)

ถ้าเป็นตัวแปรต้นทางไม่มีลิงค์จากตัวแปรอื่นเข้าไปในตัวแปรนั้น จึงเป็นอิสระจากเงื่อนไขการเกิดขึ้นของตัวแปรก่อนหน้า จึงไม่มีเงื่อนไขกับตัวแปรอื่น ๆ ดังนั้นความน่าจะเป็นของตัวแปรต้นเหตุของปัญหา ที่อยู่ในสถานะ x สามารถกำหนดตัวเลขความน่าจะเป็น ในรูปตารางความน่าจะเป็น (PT)

ถ้าหากเป็นตัวแปรผลกระทบมักเป็นเป็นตัวแปรย่อย (Child Variable) เนื่องจากมักมีตัวแปรหลักที่เชื่อมโยงอยู่ (Parent Variable) จึงจะเป็นตัวแปรที่เกิดขึ้นตามตัวแปรอื่น ๆ หรือขึ้นอยู่กับเงื่อนไขของตัวแปรก่อนหน้า ดังนั้นความน่าจะเป็นของการเกิดตัวแปรผลขึ้นอยู่กับสาเหตุของมัน ความน่าจะเป็นของตัวแปรผลกระทบที่อยู่ในสถานะที่กำหนดสามารถกำหนดโดยสถานะ (State) ที่ทราบว่าคือสถานะใดในตัวแปรหลัก และแสดงความน่าจะเป็นในตารางความน่าจะเป็นเงื่อนไข (CPT)

ในการกำหนดค่าความน่าจะเป็น มี 2 กรณีพิเศษ ถ้าเหตุการณ์สาเหตุเกิดขึ้นจะมีผลต่อเหตุการณ์ผลกระทบในลักษณะที่จะเกิดขึ้นได้อย่าแน่นอน (ความน่าจะเป็น = 1) หรือไม่เกิดขึ้นอย่างแน่นอน (ความน่าจะเป็น = 0) โดยจะเรียกทั้ง 2 กรณีนี้ว่า เป็นเหตุการณ์ที่เกิดขึ้นด้วยความเชื่อมั่นซึ่งสามารถแสดงถึงความสัมพันธ์เชิง Deterministic สมบูรณ์

2.1.3 การอนุมานด้วยโมเดลเครือข่ายแบบเบย์

2.1.3.1 หลักการเบื้องต้นของการอนุมานด้วยโมเดลเครือข่ายแบบเบย์

เมื่อกำหนดค่าความน่าจะเป็นให้กับโมเดลเรียบร้อยแล้ว ขั้นต่อไปคือการใช้ประโยชน์จากโมเดลโดยการวิเคราะห์ผลของโมเดลเครือข่ายแบบเบย์ โดยการใช้การอนุมานจากโมเดลด้วยทฤษฎีเบย์

จากโครงสร้างโมเดลเครือข่ายแบบเบย์ที่กำหนดไว้ดังรูป 2-2 สมมุติให้ ตัวแปร A เป็นตัวแปรสาเหตุเพียงตัวแปรเดียวของตัวแปร B เช่น จากตัวอย่างข้างต้น A คือ ‘การสูบบุหรี่’ และ B คือ ‘โรคหัวใจและหลอดเลือด’ การคำนวณโดยใช้ทฤษฎีเบย์ ทำได้โดย

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \qquad \textbf{(2-1)}$$

เมื่อ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$P(A)$ คือ ความน่าจะเป็นของการเกิดเหตุการณ์ A

$P(B|A)$ คือ ความน่าจะเป็นแบบมีเงื่อนไขของเหตุการณ์ B โดยกำหนดว่าเหตุการณ์ A เกิดขึ้น

สำหรับรายละเอียดการคำนวณและการอนุมานจากทฤษฎีสามารถอ่านได้เพิ่มเติมจาก Jensen & Nielsen (2007)

2.1.4 การอนุมานโมเดลเครือข่ายแบบเบย์ที่มีความซับซ้อนด้วยความเป็นอิสระอย่างมีเงื่อนไข

โดยทั่วไปแล้วโมเดลเครือข่ายแบบเบย์จะมีความซับซ้อนมากขึ้นเมื่อเป็นโครงสร้างของตัวแปรหลายตัวและทำให้มีลูกศรจำนวนมากเชื่อมโยงในโมเดล ซึ่งทำให้การอนุมานโดยใช้โมเดลเครือข่ายแบบเบย์ทำได้ยากขึ้น โดยเฉพาะอย่างยิ่งในขั้นตอนการหาการแจกแจงความเป็นไปได้ร่วม (Joint Probability Distribution) เพื่อนำมาคำนวณค่าความน่าจะเป็นของแต่ละตัวแปรที่จะเกิดขึ้น (Marginal Probability) และความน่าจะเป็นภายหลัง (Marginal Posterior Probability) ตามทฤษฎีโมเดลเครือข่ายแบบเบย์ที่นำเอาทฤษฎีความเป็นอิสระตามเงื่อนไข (Conditional Independence) มาใช้ในการอนุมานในโมเดลเครือข่ายแบบเบย์ โดยสามารถแสดงการอธิบายในรายละเอียดในหัวข้อนี้

โดยทั่วไปโมเดลเครือข่ายแบบเบย์ ที่ประกอบด้วย n ตัวแปร ที่เชื่อมโยงกันเป็นลำดับด้วยลูกศรทางเดียวจนเกิดเป็นเครือข่าย (Network) แสดงด้วย กราฟระบุทิศทาง (Directed Graphs) แสดงได้โดย $X_i \longrightarrow X_{i+1}$ สำหรับ i = 1, 2, 3, ... , n−1 โดยที่ไม่ทำให้เกิดรูปแบบโครงสร้างแบบไซเคิล (Cycle) ซึ่งจะเรียกว่า “A Directed Acyclic Graph (DAG)” ดังนั้นจากกฎ Chain Rule หรือ Markov Property ทำให้สามารถคำนวณค่า การแจกแจงความเป็นไปได้ร่วม (Joint Probability Distribution) ของตัวแปร N ตัวที่เกิดจากผลคูณของความน่าจะเป็นแบบมีเงื่อนไขที่ถูกกำหนดในโมเดลเครือข่ายแบบเบย์ โดยมีรูปทั่วไป แสดงได้ดังนี้:

$$P(X_1,...,X_n) = \prod_{i=1}^{n} P(X_i|X_1,...,X_{i-1}) \qquad \textbf{(2-2)}$$

จากตัวอย่างโครงสร้างโมเดลเครือข่ายแบบเบย์ที่กำหนดในรูป 2-1 พบว่า ตัวแปร Y เป็นตัวแปรที่ไม่มีลูกศรออกจากตัวแปรดังกล่าว ซึ่งเรียกว่า “Top Variable” และรูปแบบโครงสร้างของโมเดลที่เกิดจากความสัมพันธ์ของตัวแปร W, X , Z ที่ไม่ทำให้เกิดโครงสร้างแบบไซเคิล (Cycle) ดังนั้นด้วยกฎของ Chain Rule ทำให้สามารถคำนวณค่าความน่าจะเป็นร่วม (Joint Probability) ของตัวแปร *W,* X , Y, Z ได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$P(W,X,Y,Z) = P(Y|W,X,Z)P(X|W,Z)P(W|Z)P(Z)$$
$$= P(Y|W,X,Z)P(X|W,Z)P(W)P(Z) \quad \text{(2-3)}$$

จะเห็นได้ว่าตัวแปรที่เป็นเงื่อนไขเกิดจากการรวมเอาตัวแปรที่เป็นตัวแปร Parent Variables ของตัวแปรนั้นๆ กับตัวแปร Non-Descendent Variables เท่านั้น ซึ่งสามารถพิจารณาได้จากสมการโครงสร้างของกราฟที่สร้างขึ้น

ยกตัวอย่างเช่น การกำหนดพจน์แรกของความน่าจะเป็นแบบมีเงื่อนไขจาก Top Variable: Y โดยใช้สัญลักษณ์ $P(Y|W,X,Z)$ เนื่องจากตัวแปร W, X, Z ทั้ง 3 ตัวเป็น Non-Descendent Variables โดยจากโครงสร้างจะเห็นได้ว่าเฉพาะตัวแปร X ที่เป็นตัวแปร Parent ของ ตัวแปร Y โดยที่ W, Z เป็น Non-Descendent Variables

ดังนั้นจากเงื่อนไขที่กำหนดในสมการ (2-2) สามารถเขียนในรูปใหม่โดยแบ่งเงื่อนไขเป็น 2 ส่วน แสดงดังสมการ (2-4)

$$P(X_1,...,X_n) = \left[\prod_{i=1}^{n} P(X_i \,|\, \boldsymbol{PA}(X_i), \boldsymbol{Non_Descendent}(X_i)) \right] \quad \text{(2-4)}$$

เมื่อ

$P(X_1,...,X_n)$ คือ ความน่าจะเป็นร่วม (Joint Probability Distribution),

$\mathbf{PA}(X_i)$ คือ เซทของตัวแปร parent variables ของตัวแปร X_i,

$\mathbf{Non_Descendent}(X_i)$ คือ เซท ของตัวแปรทั้งหมด ใน DAG ที่ไม่ใช่ตัวแปรที่ไม่ใช่ $\mathbf{PA}(X_i)$ และ $\mathbf{Descendent}(X_i)$ ของตัวแปร X_i

ขั้นต่อไปเป็นการนำเสนอคุณสมบัติของโมเดลเครือข่ายแบบเบย์ที่สำคัญที่เรียกว่า “Conditional Independence” ซึ่งเป็นการอธิบายต่อเนื่องจากสมการ (2-4) ที่กำหนดให้ X_i เป็น Conditionally Independent จากเซทของตัวแปร $\mathbf{Non_Descendent}(X_i)$ โดยที่ทราบว่า เซท $\mathbf{PA}(X_i)$ คือ Parents ของตัวแปร X_i. ดังนั้นสมการ (2-4) สามารถลดรูปความซับซ้อนของกลุ่มตัวแปรเงื่อนไขลง โดยพิจารณาเฉพาะ $\mathbf{PA}(X_i)$ เป็นเงื่อนไขเฉพาะตัวแปร Parents ของตัวแปร X_i ทำให้สมการลดรูปลงเหลือสมการ (2-5)

$$P(X_1,...,X_n) = \prod_{i=1}^{n} P(X_i|\mathbf{PA}(X_i)) \quad \text{(2-5)}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตัวอย่างจากโครงสร้างโมเดลที่กำหนดสามารถหาค่าความน่าจะเป็นร่วม (Joint Probability) โดยอาศัย “Conditional Independence” ที่ใช้กำหนดความน่าจะเป็นร่วมดังสมการ (2-5) ได้ดังสมการ (2-6)

$$P(W,X,Y,Z) = P(Y|X)P(X|W,Z)P(W)P(Z) \qquad \textbf{(2-6)}$$

ลำดับถัดไปคือการคำนวณค่าความน่าจะเป็นของแต่ละเหตุการณ์ที่สามารถคำนวณได้จากการแจกแจงความน่าจะเป็นร่วม (Joint Probability Distribution) ผ่านสมการ (2-6) ซึ่งการคำนวณความน่าจะเป็นร่วม (Joint Probability) สามารถคำนวณได้จากความน่าจะเป็นของแต่ละตัวแปรที่จะเกิดขึ้น (Marginal Probability)

ตัวอย่าง 2-4 จากตัวอย่าง 2-1 เป็นความสัมพันธ์ของตัวแปรโมเดลเครือข่ายแบบเบย์ที่ประกอบด้วย 4 ตัวแปร

รูป 2-4 ตัวอย่างโมเดลเครือข่ายแบบเบย์ ที่ประกอบด้วย 4 ตัวแปร

ความน่าจะเป็นภายหลังของแต่ละตัวแปรที่จะเกิดขึ้น (Marginal Posterior Probability) ของตัวแปร X โดยกำหนด ค่าของตัวแปร Y สามารถสังเกตได้ ดังนั้น $P(X|Y)$, กำหนดดังสมการ (2-2) ได้ดังนี้

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

เมื่อ $$P(X,Y) = \sum_{W}\sum_{Z} P(W,X,Y,Z)$$

$$P(Y) = \sum_{X}\sum_{W}\sum_{Z} P(W,X,Y,Z)$$

จากตัวอย่างที่กำหนดโดยตัวแปร 4 ตัวแปร อาจต้องอาศัยกระบวนการคำนวณที่ค่อนข้างมากจึงไม่นำมาแสดงการคำนวณในที่นี้ ซึ่งผู้อ่านสามารถดูรายละเอียดได้ดังหนังสือต่างๆ เช่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Kjaerulff & Madsen, 2008; Lauritzen & Spiegelhalter, 1988) ด้วยปัญหาการคำนวณที่ยุ่งยากซับซ้อน การอนุมานด้วยโมเดลเครือข่ายแบบเบย์ จึงมีการคำนวณด้วย Algorithm และซอฟท์แวร์ต่างๆ ที่สร้างขึ้นเพื่อการศึกษาหรือเพื่อการค้า เพื่อใช้กับโมเดลที่มีขนาดใหญ่และมีความซับซ้อน Genie Software ถูกเลือกเพื่อใช้ในการสนับสนุนในการสร้างโมเดลและการวิเคราะห์ผลด้านการวิเคราะห์ความเสี่ยง (Risk Analysis)

2.1.5 Normalised likelihood (NL)

NL คืออัตราส่วนของความน่าจะเป็นที่เปลี่ยนไปของแต่ละปัจจัยเมื่อทราบสถานะของตัวแปรเป้าหมาย (Adjusted Probability) และความน่าจะเป็นที่เกิดขึ้นในปัจจุบัน (Current Probability) วัตถุประสงค์ของการคำนวณค่า NL เพื่อวิเคราะห์ความไว (Sensitivity Analysis) (Kjaerulff & Madsen, 2008) การให้เหตุผลในเชิงวินิจฉัย (Diagnostic Reasoning) จาก Top Event (Y) ไปยังตัวแปรสาเหตุแต่ละตัว (ε) ซึ่ง NL สามารถคำนวณได้จากความน่าจะเป็นภายหลังของแต่ละตัวแปรที่จะเกิดขึ้น (Marginal Posterior Probabilities): $P(\varepsilon|Y)$ ได้ดังนี้

$$NL = \frac{P(\varepsilon|Y)}{P(\varepsilon)} = \frac{P(Y,\varepsilon)/P(Y)}{P(\varepsilon)} = \frac{P(Y,\varepsilon)/P(\varepsilon)}{P(Y)} = \frac{P(Y|\varepsilon)}{P(Y)} \qquad \textbf{(2-7)}$$

เมื่อพิจารณาตัวแปรสาเหตุเพียงตัวแปรเดียว (ε), สมการ (2-7) สามารถใช้เพื่อวัดผลกระทบจากตัวแปรสาเหตุแต่ละตัวที่มีต่อตัวแปร Top Event (Y) ได้ ดังนั้น NL การเปรียบเทียบในรูปอัตราส่วนเพื่อเทียบกับความน่าจะเป็นที่เกิดขึ้นในปัจจุบัน (Current Probability) (ที่ไม่มีข้อมูลว่าเกิด Top Event หรือไม่), $P(\varepsilon)$ ผลกระทบของแต่ละเหตุการณ์ ต่อ Top Event สามารถวัดได้ด้วยอัตราส่วนระหว่าง ความน่าจะเป็นภายหลังของแต่ละตัวแปรที่จะเกิดขึ้น (Marginal Posterior Probability) (ที่ทราบว่าเกิด Top Event เกิดขึ้น), $P(\varepsilon|Y)$, และ Marginal Prior Probability (ที่ไม่มีข้อมูลว่าเกิด Top Event หรือไม่), $P(\varepsilon)$ หรือพูดได้อีกอย่างว่า NL เท่ากับ อัตราส่วนของความน่าจะเป็นที่เปลี่ยนไปของแต่ละปัจจัยเมื่อทราบว่าตัวแปรเป้าหมาย (Adjusted Probability) และความน่าจะเป็นที่เกิดขึ้นในปัจจุบัน (Current Probability) ของตัวแปรต่างๆ

"ความน่าจะเป็นที่เกิดขึ้นในปัจจุบัน" (Current Probability) vs. "ความน่าจะเป็นที่เปลี่ยนไปของแต่ละปัจจัยเมื่อทราบสถานะของตัวแปรเป้าหมาย" (Adjusted Probability)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.6 ความน่าจะเป็นภายหลังของแต่ละตัวแปร (Marginal Posterior Probability)

ความน่าจะเป็นของแต่ละปัจจัยจะมีการปรับเปลี่ยนไปตามเงื่อนไขที่กำหนด โดยใช้ทฤษฎีโมเดลเครือข่ายแบบเบย์ในการอนุมาน ซึ่งในที่นี้จะเรียกว่าความน่าจะเป็นที่ปรับแล้ว (Adjusted Probability) ดังนั้นสามารถนำค่าความน่าจะเป็นที่ปรับแล้วนำมาใช้เพื่อเปรียบเทียบสถานการณ์ต่างๆ (Scenarios) ซึ่งสามารถกำหนดสถานการณ์โดยการกำหนดให้เหตุการณ์ไม่พึงประสงค์ (Adversed Events) หรือสถานการณ์การเกิดหรือทราบเหตุการณ์ไม่พึงประสงค์ให้เกิดขึ้นพร้อมกันหลายเหตุการณ์ หรือทราบสถานะของตัวแปรเป้าหมาย

ดังนั้นเมื่อมีการคำนวณความน่าจะเป็นที่ปรับแล้ว (Adjusted Probability) เมื่อทราบสถานะของตัวแปรเป้าหมาย หรือเซทของตัวแปรเป้าหมาย ที่กำหนดโดยการกำหนดสถานการณ์

เมื่อ ε คือ เซทหรือชุดของเหตุการณ์ที่สนใจว่าเกิดขึ้นตามสถานการณ์ที่กำหนด และ

X คือ ตัวแปรที่สนใจที่ต้องการหาค่าความน่าจะเป็น ซึ่งสามารถนำเสนอในรูปความน่าจะเป็นภายหลังของแต่ละตัวแปร (Marginal Posterior Probability) ของตัวแปร X สามารถคำนวณได้ดังนี้:

$$P(X|\boldsymbol{\varepsilon}) = \frac{P(X,\boldsymbol{\varepsilon})}{P(\boldsymbol{\varepsilon})} \qquad \textbf{(2-8)}$$

ความน่าจะเป็นภายหลัง (Posterior Probability) สามารถคำนวณได้ด้วยทฤษฎีเบย์ (Bayes' Theorem) ที่คิดค้นโดย Thomas Bayes (1772 - 1761) โดยมีรูปแบบทั่วไปเป็นดังนี้

$$P(X|\varepsilon) = \frac{P(\varepsilon|X)P(X)}{P(\varepsilon)} \qquad \textbf{(2-9)}$$

เนื่องจากลักษณะที่ซับซ้อนในการคำนวณด้วยตนเองสำหรับเครือข่ายที่ซับซ้อน เราจึงไม่สามารถแสดงการคำนวณตัวอย่างได้ที่นี่ แต่สามารถใช้ซอฟท์แวร์ Genie เพื่อสนับสนุนการวิเคราะห์ อย่างไรก็ตามวิธีวิเคราะห์แบบขยายสำหรับเครือข่ายที่ซับซ้อนจะอธิบายแนวคิดในหัวข้อถัดไป

โดยทั่วไปความน่าจะเป็นภายหลัง (Posterior Probability) หมายถึงความเป็นไปได้ของแต่ละตัวแปรที่อยู่ในสถานะเฉพาะที่กำหนด เมื่อมีการระบุว่าทราบค่าสถานะ (State) ของตัวแปรที่กำหนดไว้ที่ระบุไว้ ในงานวิจัยนี้ให้ความหมายของความน่าจะเป็นภายหลัง (Posterior Probability) ว่าเป็นความน่าจะเป็นที่เปลี่ยนไป (Adjusted Probability) ของปัจจัยเสี่ยงหรือปัจจัยไม่พึงประสงค์หลังจากทราบสถานะเฉพาะของชุดตัวแปรเป้าหมาย ซึ่งก็คือตัวแปรระบุว่าเป็นโรคหรือไม่นั่นเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 การวิเคราะห์ด้วยโมเดลเครือข่ายแบบเบย์

จากการทบทวนวรรณกรรมพบว่ามีการกล่าวถึงแต่ละวิธีในการเปรียบเทียบโมเดล แต่มีน้อยมากที่กำหนดหลักเกณฑ์หรือแนวทางในการเลือกเกณฑ์ที่เหมาะสมในการประเมินประสิทธิภาพของโมเดลเครือข่ายแบบเบย์ งานวิจัยนี้ขอนำข้อเสนอของ Marcot (2012) ที่สรุปการเลือกวิธีการในการประเมินประสิทธิภาพของโมเดลตัวแบบเบย์ซึ่งพิจารณาจากวัตถุประสงค์ของการสร้างโมเดล โดยแบ่งวัตถุประสงค์ของการสร้างโมเดลเครือข่ายแบบเบย์ เป็น 5 แบบ โดยนำเสนอวิธีย่อยในแต่ละกลุ่มวัตถุประสงค์สรุปได้ดังตาราง 2-1

จากตารางที่ 2-1 เป็นกรอบการใช้ในการเลือกวิธีการประเมินโมเดลที่เหมาะสมสำหรับงานวิจัยนี้ ซึ่งก็คือ Influence analysis เพื่อให้สามารถเรียงลำดับปัจจัยสาเหตุได้อย่างเหมาะสม หากสามารถลดปัจจัยสาเหตุได้ก็จะทำให้ลดผลที่เกิดจากปัจจัยสาเหตุดังกล่าวได้อย่างเหมาะสม นอกจากนี้ Influence analysis ยังมีประโยชน์ในการประเมิน Scenarios และกำหนดขอบเขตของการจัดการเพื่อควบคุมผลกระทบจากตัวแปรที่สนใจที่ส่งผลค่าคาดหวังต่อตัวแปรเป้าหมาย (Marcot, 2012) ได้อีกด้วย การวิเคราะห์ด้วยวิธีนี้จะมีประโยชน์ในมุมมองที่ช่วยให้ผู้มีอำนาจตัดสินใจทราบถึงตัวแปรที่เป็นต้นเหตุของการเกิดโรคไม่ติดต่อที่สำคัญเพื่อกำหนดนโยบายที่เหมาะสมด้วยหลักการจากตารางประเมินความเสี่ยง (Risk Matrix) ในลำดับถัดไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 2-1 วิธีการวิเคราะห์ด้วยวิธีเครือข่ายแบบเบย์ ในวัตถุประสงค์ต่างๆ

จุดประสงค์/วิธี	รายละเอียด
1. **Metrics of model sensitivity and influence** a. **Sensitivity analysis** b. **Influence analysis**	Model sensitivity is to measure degree of which variation in PPD is explained by other variables, and essentially depicts the underlying probability structure of a model given prior probability distribution. Whereas influence analysis aim to evaluate effects on PPDs from selected input variables set to best or worst case scenario values. Conducting influence runs can help reveal the degree to which individual or sets of input variables could affect outcome probability.
2. **Metrics of model complexity**	Numbers of variables and numbers of links
3. **Metrics of BN model prediction performance** a. **Error rates and confusion tables** b. **Weighted confusion error rates** c. **ROC curves and AUC** d. **K-fold cross validation** e. **Spherical payoff** f. **Schwarz' Bayesian information criterion** g. **True skill statistic and Cohen's kappa**	Evaluate performance accuracy of BN model prediction
4. **Metrics of uncertainty in posterior probability distribution** a. **Bayesian credible intervals** b. **Posterior probability certainty index** c. **Certainty envelope** d. **Inequality of posterior probability distribution**	Degree of uncertainty in BN model outcomes, that is, the dispersion of PPD values.
5. **Metrics for comparing alternative posterior probability distributions**	To compare results of running BN models among alternative scenarios is to directly compare PPD with existing statistical tests.

ที่มา: ปรับปรุงจาก (Marcot, 2012)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 การตรวจสอบความถูกต้องของตัวแบบในการพยากรณ์

ในการพยากรณ์การเกิดโรคด้วยตัวแบบต่างๆ ต้องตรวจสอบความถูกต้องในการพยากรณ์และค่าที่แท้จริงโดยเปรียบเทียบจากค่าสถิติดังต่อไปนี้

2.3.1 คอนฟูชั่น แมททริก (Confusion Matrix)

คอนฟูชั่น แมททริกเป็นตารางที่มักใช้ในการอธิบายถึงประสิทธิภาพของแบบจำแนก (Classifier) ในชุดของข้อมูลที่ทราบค่าจริง คอนฟูชั่น แมททริกแบบง่ายๆ ที่ใช้เพื่อจำแนกกลุ่ม 2 กลุ่ม แสดงดังตาราง 2-2

ตาราง 2-2 คอนฟูชั่น แมททริก (**Confusion Matrix**) ในการจำแนก 2 กลุ่ม

<table>
<tr><th rowspan="2">ค่าจริง (Actual)</th><th colspan="2">ค่าพยากรณ์ (Predictive)</th><th colspan="2" rowspan="2">ค่าวัดปะสิทธิภาพ</th></tr>
<tr><th>Positive</th><th>Negative</th></tr>
<tr><td>Positive</td><td>ผลบวกจริง (TP)</td><td>ผลลบลวง (FN)</td><td>อัตราความถูกต้องเชิงบวก (TPR) = ความไว (Sensitivity) = ความระลึก (Recall)
$TPR = \frac{TP}{TP + FN}$</td><td>$FNR = \frac{FN}{TP + FN}$</td></tr>
<tr><td>Negative</td><td>ผลบวกลวง (FP)</td><td>ผลลบ
จริง (TN)</td><td>อัตราความถูกต้องเชิงลบ (TNR) = ความจำเพาะ (Specificity) = TN / (FP + TN)</td><td>$FPR = \frac{FP}{FP + TN}$</td></tr>
<tr><td>ค่าวัดปะสิทธิภาพ</td><td>ค่าความแม่นยำ (Precision)
$P = \frac{TP}{TP + FP}$</td><td>ค่าทำนายเมื่อผลเป็นลบ (NPV) = TN / (FN + TN)</td><td colspan="2">Accuracy = (TP+TN)/(TP+FN+FP+TN)</td></tr>
</table>

เมื่อ True Positives (TP): จำนวนข้อมูลที่จำแนกถูกว่าเป็นโรคและเป็นโรคจริง

True Negatives (TN): จำนวนข้อมูลที่จำแนกถูกว่าไม่เป็นโรคและไม่เป็นโรคจริง

False Positives (FP): จำนวนข้อมูลที่จำแนกผิดว่าเป็นโรคแต่จริงๆ แล้วไม่เป็นโรคจริง

False Negatives (FN): จำนวนข้อมูลที่จำแนกผิดว่าไม่เป็นโรคแต่จริงๆ แล้วเป็นโรคจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางคอนฟูชั่น แมททริก สามารถนำมาคำนวณค่าเพื่อตรวจสอบความสามารถในการพยากรณ์เพื่อการจำแนกกลุ่มได้หลายเกณฑ์ ซึ่งในงานวิจัยนี้

1. *อัตราความถูกต้องเชิงบวก (True Positive Rate : TPR)*

$$TPR = \frac{TP}{TP + FN} \quad \textbf{(2-10)}$$

TPR = จำนวนข้อมูลที่พยากรณ์ถูกว่าอยู่กลุ่มที่สนใจ / จำนวนข้อมูลที่แท้จริงในกลุ่มที่สนใจ

2. *อัตราความถูกต้องเชิงลบ (True Negative Rate : TNR)*

$$TNR = \frac{TN}{FP + TN} \quad \textbf{(2-11)}$$

TNR = จำนวนข้อมูลที่พยากรณ์ถูกว่าอยู่กลุ่มที่ไม่สนใจ / จำนวนข้อมูลที่แท้จริงในกลุ่มที่ไม่สนใจ

3. *อัตราความผิดพลาดเชิงบวก (Fault Positive Rate : FPR)*

$$FPR = \frac{FP}{FP + TN} \quad \textbf{(2-12)}$$

FPR = จำนวนข้อมูลที่พยากรณ์ผิดว่าอยู่กลุ่มที่สนใจ / จำนวนข้อมูลที่แท้จริงในกลุ่มที่ไม่สนใจ

4. *อัตราความผิดพลาดเชิงลบ (Fault Negative Rate : FNR)*

$$FNR = \frac{FN}{TP + FN} \quad \textbf{(2-13)}$$

FNR = จำนวนข้อมูลที่พยากรณ์ผิดว่าอยู่กลุ่มที่ไม่สนใจ / จำนวนข้อมูลที่แท้จริงในกลุ่มที่สนใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. *ค่าความแม่นยำ (Precision)*

$$Precision = \frac{TP}{TP + FP} \quad \textbf{(2-14)}$$

Precision = จำนวนข้อมูลที่พยากรณ์ถูกว่าอยู่กลุ่มที่สนใจ / จำนวนข้อมูลที่ทำนายว่าอยู่ในกลุ่มที่สนใจ

6. *ความถูกต้อง (Accuracy)*

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \textbf{(2-15)}$$

Accuracy = จำนวนข้อมูลที่พยากรณ์ถูกว่าอยู่กลุ่มที่สนใจ หรือกลุ่มอื่น / จำนวนข้อมูลที่ทั้งหมด

7. *ความไว (Sensitivity) หรือ ความระลึก (Recall)*

$$Sensitivity = \frac{TP}{TP + FN} = TPR \quad \textbf{(2-16)}$$

8. *ความจำเพาะ (Specificity)*

$$Specificity = \frac{TN}{FP + TN} = TNR \quad \textbf{(2-17)}$$

2.3.2 ค่า F1

ในกรณีที่ค่าความสามารถในการจำแนกของค่าต่างๆ ข้างต้น มีค่าไม่เป็นไปในทิศทางเดียวกัน มีความจำเป็นที่ต้องหาค่ากลางตัวใหม่ที่เป็นค่าที่ใช้ตัดสินว่าโมเดลใดมีความสามารถในการพยากรณ์ดีกว่ากัน ค่าที่มักมีมีผลขัดแย้งกันคือ ค่าความไว (Sensitivity) และความจำเพาะ (Specificity) เป็นค่าวัดทางสถิติที่ใช้ประเมินประสิทธิภาพของการทดสอบที่ให้ผลเป็น 2 ส่วน (เช่น เป็นบวกและลบ) โดย แต่ละค่ามีรายละเอียดดังนี้

ความไว (Sensitivity) คือ สัดส่วนของผลบวกที่เป็นจริงสำหรับภาวะนั้น ๆ (เช่น สัดส่วนของการตรวจพบโรคในผู้ที่ป่วยจริง) มีคำอื่น ๆ ที่มีความหมายเช่นเดียวกัน เช่น อัตราผลบวกจริง (TPR : True Positive Rate), ความระลึก (Recall), ความน่าจะเป็นในการตรวจพบ (Probability of Detection) ซึ่งมักใช้แตกต่างกันในสาขาต่าง ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความจำเพาะ (Specificity) คือ สัดส่วนของผลลบที่เป็นจริงสำหรับภาวะนั้น ๆ (เช่น สัดส่วนของการตรวจไม่พบโรคในผู้ที่ไม่ป่วย) มีคำอื่น ๆ รวมทั้ง อัตราผลลบจริง (TNR : True Negative Rate) โดยความจำเพาะจึงมีประโยชน์ในการยืนยันภาวะที่มี โดยป้องกันการเกิด อัตราความผิดพลาดเชิงบวก (False Positive) เพราะว่าการทดสอบยิ่งจำเพาะเท่าไร โอกาสการได้ผลบวก (เช่น การพบว่ามีโรค) ที่ไม่เป็นจริง (เช่น บุคคลจริง ๆ ไม่มีโรค) ก็น้อยลงเท่านั้น และดังนั้น ถ้าความจำเพาะอยู่ที่ 100% โอกาสได้ผลบวกปลอมก็อยู่ที่ 0%

ความไว (Sensitivity) จึงมีประโยชน์ในการวินิจฉัยแยกกันความผิดพลาดเชิงลบ (Fault Negative) เพราะว่าการทดสอบยิ่งไวเท่าไร โอกาสการได้ผลลบ (เช่น การพบว่าไม่มีโรค) ที่ไม่เป็นจริง (เช่น บุคคลจริง ๆ มีโรค) ก็น้อยลงเท่านั้น และดังนั้น ถ้าความไวอยู่ที่ 100% โอกาสได้ผลความผิดพลาดเชิงลบ ก็อยู่ที่ 0%

F1 คำนวณจากค่าเฉลี่ย ฮาร์โมนิคระหว่างค่าความแม่นยำ (Precision) ดังสมการ (2.14) และความระลึก (Recall) ดังสมการ (2-16) ใช้เพื่อเปรียบเทียบความสามารถในการแบ่งกลุ่ม และมักใช้เป็นตัวแปรชี้ขาดว่าโมเดลใดเหมาะสมมากกว่ากันเมื่อมีค่าทั้ง 2 ค่าดังกล่าวไม่เป็นไปในทิศทางเดียวกัน แสดงได้ดังสูตรนี้

$$F1 = 2 * \left(\frac{Precision * Recall}{Precision + Recall} \right) \quad \textbf{(2-18)}$$

2.3.3 เส้นโค้ง ROC และ AUC

ความถูกต้องของการพยากรณ์เป็นปัจจัยหลักในการตัดสินใจเลือกโมเดลที่เหมาะสมในโมเดลเพื่อการพยากรณ์ โดยความสามารถในการทำนายสามารถประเมินจากเส้นโค้ง ROC โดยเส้นโค้ง ROC แสดงถึงความสัมพันธ์ระหว่างเปอร์เซ็นต์ความถูกต้องเชิงบวก (True Positive) และเปอร์เซ็นต์ของความผิดพลาดเชิงบวก (Fault Positive) นอกจากนี้ยังใช้พื้นที่ใต้เส้นโค้ง ROC ที่เรียกว่า AUC โดยใช้เพื่อแสดงประสิทธิภาพโดยรวมในการจำแนกประเภท ค่า AUC อยู่ระหว่าง 0 ถึง 1 ถ้าค่า AUC เท่ากับ 1 แสดงว่ามีประสิทธิภาพที่ดีโดยไม่มีข้อผิดพลาด ในทางกลับกันเมื่อ AUC มีขนาดเล็กมาก และมีค่าใกล้เคียงกับ 0 ตัวแบบมักมีการคาดการณ์ที่ไม่ถูกต้อง เมื่อค่า AUC ใกล้เคียง 0.5 โมเดลสามารถมีความสามารถแบบสุ่ม (Marcot, 2012)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4 วรรณกรรมที่เกี่ยวข้อง (Literature Review)

2.4.1 การใช้โมเดลเครือข่ายแบบเบย์กับข้อมูลความชุกการเกิดโรคในระดับประชากร

ผลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกายครั้งที่ 1- 4 ได้มีการจัดทำในรูปแบบรายงาน ซึ่งสามารถ download ได้จากเว็บไซท์ของ 'สำนักงานพัฒนาระบบข้อมูลข่าวสารสุขภาพ' ข้อมูลจากการสำรวจยังได้นำไปใช้โดยหน่วยงานที่เกี่ยวข้องเพื่อรายงานการเฝ้าระวังโรค (อมรา ทองหงษ์ et al., 2556) หรือรายงานประจำปี (สำนักโรคไม่ติดต่อ กรมควบคุมโรค, 2556) รวมทั้งนำเสนอในรูปบทความทางวิชาการระดับนานาชาติ (ยกตัวอย่างเช่น Aekplakorn et al., 2007; Danaei et al., 2011)

จากการทบทวนวรรณกรรมยังพบว่าวิธีโมเดลเครือข่ายแบบเบย์เคยนำมาใช้เพื่อวินิจฉัยหรือพยากรณ์การเกิดโรคต่างๆ จากฐานข้อมูลผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาล เช่น การพยากรณ์ความเสี่ยงการเป็นโรคมะเร็งเต้านม (Burnside et al., 2006) การพยากรณ์การเกิดโรคโลหิตจาง (Sebastiani et al., 2007) การพยากรณ์การเกิดโรคเบาหวานประเภทที่ 2 (Yang Guo, Guohua Bai, & Yan Hu, 2012) หรือการประเมินความเสี่ยงต่อสุขภาพจากมลพิษทางอากาศ (Liu, Lu, Chen, & Shen, 2011) ผู้วิจัยพบว่าการศึกษาภายใต้ฐานข้อมูลผู้ป่วยที่มีการบันทึกตามสถานพยาบาลต่างๆ มีข้อจำกัดเพราะไม่สามารถอนุมานไปสู่ความชุกระดับประชากรผู้ที่เป็นโรคแต่ยังไม่ทราบหรือไม่ได้เข้าสู่ระบบของการรักษา คนกลุ่มนี้เป็นกลุ่มที่มีความสำคัญเพราะพวกเขายังไม่มีการปรับพฤติกรรมเสี่ยงเนื่องจากยังไม่ทราบว่าตนเองเป็นโรค

เมื่อมุ่งทบทวนวรรณกรรมเฉพาะที่นำฐานข้อมูลระดับประชากรที่ได้จากการสุ่มตัวอย่างมาจากประชากรระดับประเทศเพื่อมาสร้างโมเดลเครือข่ายแบบเบย์พบว่า โมเดลดังกล่าวไม่รวมปัจจัยความเสี่ยงจากความแตกต่างด้านลักษณะประชากรที่สำคัญ เช่น การศึกษาของ Fellaji และคณะ (2014) หรือไม่นำตัวบ่งชี้ด้านพฤติกรรมบางอย่าง เช่นพฤติกรรมการบริโภค เข้ามาเป็นส่วนหนึ่งในการสร้างโมเดลเช่น การสร้างโมเดลสำหรับพยากรณ์การเกิดโรคหัวใจและหลอดเลือด (Cardiovascular) โดย Atoui และคณะ (2006) หรือ Twardy และคณะ (2005) ซึ่งอาจไม่สามารถระบุสาเหตุที่แท้จริงของการเกิดโรคต่างๆ ที่มาจากปัจจัยความเสี่ยงทางด้านพฤติกรรมได้

นอกจากนี้ยังพบว่าข้อมูลจากการสำรวจสุขภาพของประชาชนไทยโดยการตรวจร่างกายไม่เคยถูกนำมาใช้เพื่อสร้างโมเดลแสดงความสัมพันธ์ระหว่างปัจจัยเสี่ยงที่เกี่ยวข้องกับโรคไม่ติดต่อเรื้อรังทั้ง 2 โรคดังกล่าวด้วยวิธีโมเดลเครือข่ายแบบเบย์ โดยสรุปแล้วผลจากงานวิจัยในครั้งนี้จะทำให้ได้โมเดลที่แสดงความสัมพันธ์ของสาเหตุหรือปัจจัยเสี่ยงของทั้ง 2 โรคที่ครอบคลุมปัจจัยเสี่ยงทางชีวภาพ (Biomarker) ตัวบ่งชี้ด้านพฤติกรรม (Behavioural Marker) และตัวบ่งชี้ด้านลักษณะประชากร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Characteristic Marker) จากฐานข้อมูลระดับประชากรของประเทศไทย (การสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย ครั้งที่ 4)

2.4.2 การใช้โมเดลเครือข่ายแบบเบย์ในการวิเคราะห์การเกิดโรค

จากการสำรวจงานวิจัยเกี่ยวกับโมเดลที่ช่วยในการตัดสินใจวินิจฉัยโรคพบว่าโมเดลเครือข่ายแบบเบย์เป็นหนึ่งในวิธีที่สามารถนำมาใช้และมีแนวโน้มเพิ่มมากขึ้น (Wagholikar, Sundararajan, & Deshpande, 2012) ตามที่ผู้วิจัยนำเสนอรูปแบบการวิเคราะห์สำหรับการวิจัยนี้ ตาม 3 วิธี ดังกล่าว ผู้วิจัยได้ใช้เพื่อจำแนกตัวอย่างงานวิจัยที่ได้เคยทำมาแล้ว แสดงได้ดังตาราง 2-3

ตาราง 2-3 สรุปงานวิจัยที่เกี่ยวข้องจำแนกตามจุดประสงค์ของการวิเคราะห์ด้วยโมเดลเครือข่ายแบบเบย์

พยากรณ์การเกิดโรค	**จัดลำดับปัจจัยเสี่ยงที่อาจก่อให้เกิดโรคไม่ติดต่อที่สำคัญ**	**ประมาณค่าผลกระทบของการเกิดโรคเมื่อกำหนดสถานการณ์ให้ความเสี่ยงหลายปัจจัยเกิดขึ้นพร้อมกัน**
พยากรณ์การเกิดเบาหวานชนิดที่ 2 (Yang Guo et al., 2012)	หาสาเหตุเบื้องต้นจากการกำหนด Pathway จากการเกิดโรคหัวใจและหลอดเลือด (Long, Naimi, Criscitiello, & Larsen, 1989)	วิเคราะห์สาเหตุการเกิดโรคไตโดยกำหนดเซทของปัจจัยเสี่ยงที่เกิดขึ้น (Fellaji et al., 2014)
พยากรณ์การเกิดโรคหัวใจและหลอดเลือด (Atoui et al., 2006)	-	ใช้การวิเคราะห์สถานการณ์เพื่อนำมาใช้กับ cost-effective analysis โดยนำมาใช้เพื่อประเมินความเสี่ยงด้านสุขภาพ (Liu et al., 2011)
พยากรณ์หรือประมาณค่าความน่าจะเป็นในการเกิดโรคทางระบาดวิทยาโดยยกตัวอย่างโรคท้องเสีย (Diarrhea) (Nguefack-Tsague, 2011)	-	-
พยากรณ์ค่า Severity score เพื่อตรวจหาผู้ที่มีภาวะเสี่ยงการเป็นโรค sickle cell (Sebastiani et al., 2007)	-	-

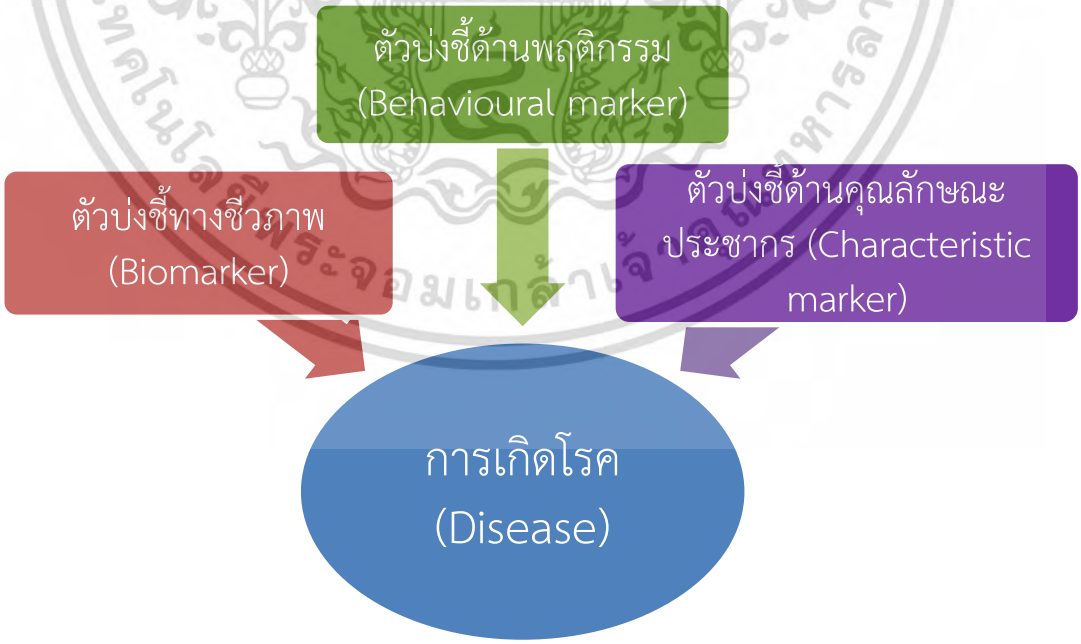
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3 วิธีการและแผนการดำเนินงานวิจัย

ในบทนี้จะนำเสนอแนวทางการดำเนินการวิจัยเพื่อสร้างโมเดลเครือข่ายแบบเบย์พร้อมทั้งแนวทางการวิเคราะห์เพื่อสนับสนุนการวิเคราะห์ความเสี่ยงในการเกิดโรคไม่ติดต่อเรื้อรังซึ่งคาดว่าจะใช้เป็นแนวทางในการวิเคราะห์ในโรคอื่นๆ ต่อไป

3.1 กรอบแนวความคิด

จากการทบทวนวรรณกรรมสามารถนำมาสู่การกำหนดปัจจัยหลักที่ส่งผลต่อสถานะทางสุขภาพของการเกิดโรคไม่ติดต่อเรื้อรัง โดยสามารถกำหนดรูปแบบความสัมพันธ์จากสาเหตุหลัก 3 ตัวบ่งชี้ ได้แก่ ตัวบ่งชี้ด้านคุณลักษณะประชากร (Characteristic marker) ตัวบ่งชี้ด้านชีวภาพ (Biomarker) และตัวบ่งชี้ด้านพฤติกรรม (Behavioural marker) จึงนำมาใช้เป็นตัวกำหนดกรอบแนวความคิดการวิจัยเพื่อนำมาเป็นกรอบในการกำหนดโมเดลเครือข่ายแบบเบย์ที่เหมาะสมต่อไป ซึ่งแสดงได้ดังรูป 3-1



รูป 3-1 กรอบการศึกษาแสดงความสัมพันธ์ของปัจจัยต่างๆ ที่ส่งผลต่อการเกิดโรค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 วิธีดำเนินการวิจัย

การดำเนินงานวิจัยเน้นการนำข้อมูลทุติยภูมิที่มีการเก็บรวมรวมไว้จากการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกายครั้งที่ 4 พร้อมด้วยข้อมูลปฐมภูมิจากการสัมภาษณ์แพทย์ผู้เชี่ยวชาญในสาขาโรคเบาหวาน และความดันโลหิตสูง (ที่ได้ระบุไว้ในขอบเขตการศึกษา) เพื่อให้แน่ใจว่าโมเดลที่สร้างขึ้นมีความถูกต้องและเหมาะสมตามหลักฐานทางการแพทย์โดยมีขั้นตอนในการสร้างโมเดลดังตาราง 3-1

ข้อมูลที่รวบรวมมาได้ในแต่ละขั้นตอนจะนำมาวิเคราะห์ที่ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ด้วยโปรแกรมพื้นฐานทางสถิติที่มีอยู่ และโปรแกรมเฉพาะทางเพื่อสร้างโมเดลเครือข่ายแบบเบย์ซึ่งเน้นโปรแกรมที่ไม่เสียค่าใช้จ่ายเพื่อลดปัญหาการเข้าถึงโมเดลโดยผู้ใช้เพื่อลดเงื่อนไขการเข้าถึงโปรแกรมและโมเดลหลังจากโครงการวิจัยเสร็จสิ้นลง

ตาราง 3-1 สรุปขั้นตอนการสร้างโมเดลเครือข่ายแบบเบย์จากข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย

ขั้นตอนการสร้างโมเดล	วิธีการ
1. กำหนดโครงสร้างของโมเดลจากรายงานการสำรวจสุขภาพของประชาชนไทยโดยการตรวจร่างกาย ครั้งที่ 4 และข้อมูลที่มีการจัดเก็บจากการสำรวจดังกล่าวและจากการทบทวนวรรณกรรม	โดยผู้วิจัย
2. ตรวจสอบความถูกต้องของโมเดลกับแพทย์ผู้เชี่ยวชาญเฉพาะโรค (โมเดลสามารถปรับเปลี่ยนแก้ไขโครงสร้างตามคำแนะนำของแพทย์ผู้เชี่ยวชาญ)	- สัมภาษณ์แพทย์ผู้เชี่ยวชาญ *4* ท่าน - สัมภาษณ์ Decision maker (พบผู้เชี่ยวชาญต่อท่าน 2 ครั้ง)
3. กำหนดค่าความน่าจะเป็นของแต่ละปัจจัยตามโครงสร้างโมเดล ในข้อ 2 จากฐานข้อมูลที่ได้จากการสำรวจฯ	ติดต่อสำนักงานการสำรวจสภาวะสุขภาพของประชาชนไทย (สสท.) เพื่อขอใช้ข้อมูลจากฐานข้อมูล
4. นำตัวเลขที่ได้ป้อนเข้าสู่ ซอฟท์แวร์ที่ใช้สำหรับโมเดลเครือข่ายแบบเบย์	โดยผู้วิจัย
5. ตรวจสอบความถูกต้องหลังจากป้อนข้อมูล	โดยผู้วิจัย
6. ใช้โมเดลที่ป้อนข้อมูลความน่าจะเป็นเรียบร้อยแล้วมาทำการวิเคราะห์	โดยผู้วิจัย
7. ตรวจสอบความถูกต้องสมบูรณ์จากผลการวิเคราะห์ที่ได้จากโมเดลกับแพทย์ผู้เชี่ยวชาญ	- สัมภาษณ์แพทย์ผู้เชี่ยวชาญ *4* ท่าน - สัมภาษณ์ Decision maker (พบผู้เชี่ยวชาญต่อท่าน 2 ครั้ง)
8. *ปรับปรุงแก้ไขโมเดลตามคำแนะนำของผู้เชี่ยวชาญก่อนนำเสนอรายงานผลการวิเคราะห์ไปยังหน่วยงานที่เกี่ยวข้อง*	โดยผู้วิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อนึ่งขั้นตอนการสร้างโมเดลได้ผ่านความเห็นชอบจากผู้เชี่ยวชาญ ทำให้ทราบถึงความเป็นไปได้ในการใช้กระบวนการสร้างโมเดล และได้นำกระบวนการดังกล่าวไปใช้ในขั้นตอนการสร้างโมเดลร่วมกัน นอกจากนี้ผู้วิจัยยังได้รับความร่วมมือจากสำนักงานการสำรวจสภาวะสุขภาพของประชาชนไทย (สสท.) ในการขออนุเคราะห์ข้อมูลจากฐานข้อมูลการสำรวจสุขภาพของประชาชนไทย

3.3 การวิเคราะห์จากโมเดลเครือข่ายแบบเบย์

การใช้ความน่าจะเป็นในการให้เหตุผลในเชิงถ่ายทอด (Propagate) ความน่าจะเป็นภายใต้แนวคิดโมเดลเครือข่ายแบบเบย์ เพื่อให้ผลลัพธ์จากกระบวนการอนุมานของโมเดลเครือข่ายแบบเบย์เพื่อสนับสนุนการวิเคราะห์ความเสี่ยงของโรคไม่ติดต่อเรื้อรัง เพื่อแสดงผลกระทบจากรูปแบบถูกกำหนดตามความต้องการของผู้ใช้ โดยพบว่า Sodhi and Tang (2012) ใช้โมเดลเครือข่ายแบบเบย์เพื่อสนับสนุนด้านการวิเคราะห์ความเสี่ยงเพื่อช่วยให้ฝ่ายบริหารมุ่งเน้นจัดการเฉพาะด้านและสนับสนุนการจัดสรรทรัพยากรที่เหมาะสมเพื่อใช้ในการบริหารความเสี่ยงเพราะทุกองค์กรมีงบประมาณที่จำกัด สำหรับงานวิจัยนี้ผู้วิจัยขอเสนอแนวทางการวิเคราะห์จากโมเดลเพื่อวัตถุประสงค์ 2 ประการ ตามแนวทางผลลัพธ์ที่คาดจะได้รับจากโมเดลดังนี้

1. วิเคราะห์ปัจจัยเสี่ยงหรือตัวบ่งชี้ความเสี่ยงที่สำคัญต่อการเกิดโรคไม่ติดต่อที่สำคัญของคนไทยจากข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกายจากโมเดลเครือข่ายแบบเบย์ (Bayesian Network Model) ด้วยการ
    a. จัดลำดับปัจจัยเสี่ยงที่อาจก่อให้เกิดโรคไม่ติดต่อที่สำคัญแต่ละโรคด้วยการอนุมานแบบ Diagnostic Inference
    b. การจัดกลุ่มความเสี่ยงเพื่อกำหนดนโยบายที่เหมาะสมของความเสี่ยงแต่ละกลุ่ม
2. พยากรณ์ความชุกของโรคและสาเหตุการเกิดโรคของคนไทยในปัจจุบันด้วยการอนุมานแบบ Prognostic inference

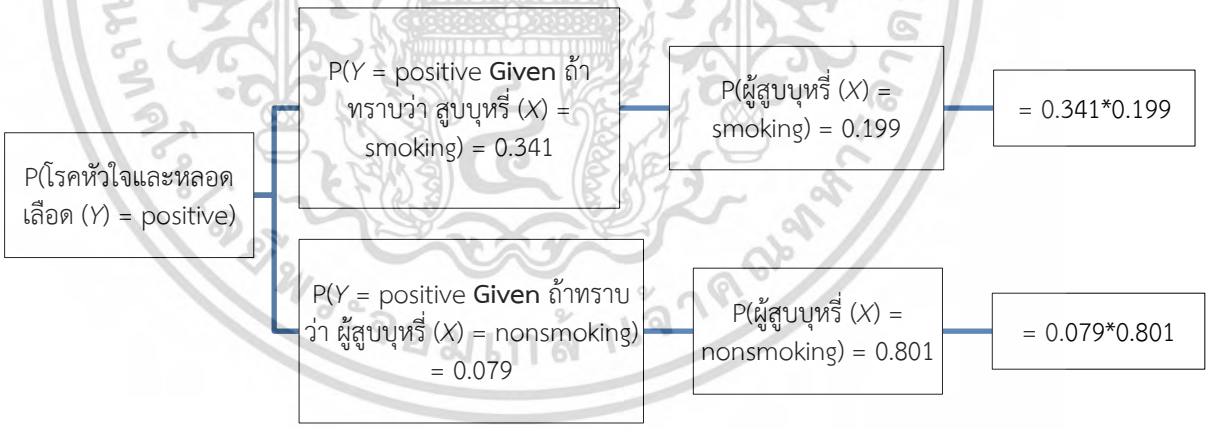
เนื่องจากปัญหาการเกิดโรคมีสาเหตุที่เกิดขึ้นได้มากมายดังนั้นเมื่อนำมาอธิบายในรูปแบบความสัมพันธ์ด้วยกราฟจึงอาจมีความซับซ้อนและเรียกกราฟดังกล่าวว่าเครือข่าย (Network) และยิ่งเครือข่ายของโมเดลมีขนาดใหญ่เท่าไหร่ ความต้องการใช้ข้อมูลความน่าจะเป็นยิ่งมีจำนวนมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.1 การอนุมานความน่าจะเป็นของการเกิดโอกาสที่ไม่พึงประสงค์

การสร้างโมเดลเครือข่ายแบบเบย์เป็นการกำหนดความสัมพันธ์ของเหตุการณ์ไม่พึงประสงค์ที่อาจก่อให้เกิดผลต่อตัวแปรผลลัพธ์ที่สนใจทั้งที่เป็นสาเหตุทางตรง (Direct Effect) และทางอ้อม (Indirect Effect) โดยโมเดลจะแสดงลิงค์เฉพาะตัวแปรที่มีความสัมพันธ์ทางตรงเท่านั้น ซึ่งในการศึกษาครั้งนี้จะเป็นการแสดงความสัมพันธ์โครงสร้างเชิงเหตุและผล (Cause-Effect Diagram) หากตัวแปรที่เป็นตัวแปรสาเหตุย่อย (Child Variable) ในขั้นตอนการกำหนดค่าความน่าจะเป็นถูกกำหนดโดยตาราง CPT ซึ่งไม่ใช่การกำหนดค่าความน่าจะเป็นของแต่ละสถานะ (State) ของตัวแปรโดยตรง ดังนั้นหากต้องการหาค่าความน่าจะเป็นของการเกิดเหตุการณ์ไม่พึงประสงค์ต่างๆ ที่อยู่ในรูปตัวแปรสาเหตุย่อย (Child Variable) หรือตัวแปรผลกระทบ (Effect Variable) ในโมเดลที่ไม่ใช้ตัวแปรต้นเหตุของปัญหา (Root Cause Variable) ต้องนำทฤษฎีการอนุมานด้วยโมเดลเครือข่ายแบบเบย์มาใช้ในการอธิบายการคำนวณ

ตัวอย่าง 3-1 จากตัวอย่างข้างต้นการกำหนดค่าความน่าจะเป็นของเหตุการณ์หรือตัวแปร ‘การสูบบุหรี่’ ที่ปรากฏใน (Probability Table: PT) แต่ค่าความน่าจะเป็นของการเกิด ‘โรคหัวใจและหลอดเลือด’ ไม่สามารถกำหนดลงในในโมเดลได้โดยตรงในตาราง CPT ซึ่งต้องอาศัยหลักการคำนวณจาก 0.341*0.199 + 0.079*0.801 = 0.131138, ซึ่งแสดงได้ดังรูป 3-2



รูป 3-2 การคำนวณโอกาสการเกิด ‘โรคหัวใจและหลอดเลือด’ (Y)

รายละเอียดแสดงการคำนวณสามารถแสดงได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$P(Y = \text{positive}) = \sum_x P\left(Y = \text{positive}\middle|X\right)P(X)$$
$$= P\left(Y = \text{positive}\middle|X = \text{smoking}\right) * P\left(X = \text{smoking}\right)$$
$$+ P\left(Y = \text{positive}\middle|X = \text{nonsmoking}\right) * P\left(X = \text{nonsmoking}\right)$$
$$= 0.341 * 0.199 + 0.079 * 0.801 = 0.067859 + 0.063279 = 0.131138$$

ความน่าจะเป็นที่จะไม่เป็นโรคหัวใจและหลอดเลือด คือ:

$$P(Y = \text{negative}) = 1 - P\left(Y = \text{positive}\right) = 1 - 0.131138 = 0.868862$$

โดยทั่วไปความน่าจะเป็นส่วนสำคัญของตัวแปร (ทั้งสาเหตุหรือผลกระทบตัวแปร) คือความน่าจะเป็นของการเกิดขึ้นสำหรับแต่ละสถานะ (State) ของตัวแปรโดยเฉพาะ ซึ่งเรียกว่า “Marginal Prior” โดยในรายงานนี้จะเรียกว่า “ความน่าจะเป็นในปัจจุบัน” ของเหตุการณ์ไม่พึงประสงค์ที่สนใจ ความน่าจะเป็นดังกล่าวใช้เพื่ออธิบายความเป็นไปได้ของเหตุการณ์ไม่พึงประสงค์ที่จะเกิดขึ้นของแต่ละตัวแปรในโมเดล

อนึ่งเฉพาะตัวแปรปัจจัยที่เป็นต้นเหตุของปัญหา (Root Cause) ซึ่งเป็นตัวแปรต้นกำเนิดของการเกิดโรคในโมเดล จะมีความน่าจะเป็นเท่ากับความน่าจะเป็นที่ป้อนข้อมูล (Input) ในตาราง PT แต่สำหรับตัวแปรตัวแปรที่ไม่ใช่ตัวแปรต้นเหตุ หรือเรียกว่า ‘Child Variable’ หรือผลลัพธ์ที่มีผลต่อความน่าจะเป็นของแต่ละตัวแปรที่จะเกิดขึ้น (Marginal Probability) ที่ไม่ใช่ข้อมูลนำเข้า (Input) โดยตรงและจะได้จากการคำนวณโดยสมการ (3-1)

ตัวอย่าง 3-1 แสดงความสัมพันธ์ของตัวแปร 2 ตัวแปรเมื่อตัวแปร X เป็นสาเหตุของตัวแปร Y ดังรูป 3-3

รูป 3-3 ตัวอย่างโมเดลเครือข่ายแบบเบย์ซึ่งแสดงความสัมพันธ์ของตัวแปร 1 คู่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การคำนวณโอกาสของการเกิด Y สามารถคำนวณได้จากสมการ (3-1)

$$P(Y) = \sum_{x} P(Y|X)P(X) \qquad \textbf{(3-1)}$$

3.3.2 จัดลำดับปัจจัยเสี่ยงที่อาจก่อให้เกิดโรคไม่ติดต่อที่สำคัญ

พิจารณาตัวอย่างง่ายๆของตัวแปรสองตัวที่ X เป็นสาเหตุของ Y และ Y เป็นเหตุการณ์ด้านบน (รูป 3-3) จากนั้นความน่าจะเป็นหลังด้านหลังของ X เมื่อกำหนดว่า Y ได้เกิดขึ้น สามารถคำนวณได้โดย สมการ (3-2)

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \qquad \textbf{(3-2)}$$

ตัวอย่าง 3-2 อ้างถึงตัวอย่าง 2-3 หากมีการสังเกตว่ามีผู้ป่วยโรคหัวใจและหลอดเลือด โอกาสที่จะเกิดจากสาเหตุของการที่เขาเป็นผู้สูบบุหรี่เป็นเท่าใด

ในงานวิจัยนี้จะใช้ความน่าจะเป็นที่เปลี่ยนไปของแต่ละปัจจัยเมื่อทราบว่าเป็นผู้ป่วยโรคเบาหวาน (Adjusted Probability) ของตัวแปรการเป็นผู้สูบบุหรี่ (X) เมื่อทราบว่าเขาเป็นผู้ป่วยโรคหัวใจและหลอดเลือด (Y) คำนวณได้จาก

$$P(X = smoker|Y = positive) = \frac{P(Y = positive|X = smoker)P(X = smoker)}{P(Y = positive)}$$

เมื่อกำหนดให้ $P(X)$ แทน Prior Probability และ $P(Y|X)$ แทน Conditional Probability ที่แสดงการกำหนดค่าในตัวอย่าง 3-2 ในรูปข้อมูลที่อยู่ในตาราง PT และ CPT เมื่อกำหนดค่าความน่าจะเป็นของแต่ละตัวแปรที่จะเกิดขึ้น (Marginal Probability) ของ $Y =$ positive แสดงได้ว่า

$$P(X = smoker|Y = positive) = \frac{P(Y = positive|X = smoker)P(X = smoker)}{P(Y = positive)}$$

$$= \frac{0.341 * 0.199}{0.341 * 0.199 + 0.079 * 0.801} = \frac{0.067859}{0.131138} = 0.5175$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้นเมื่อเราสังเกตว่าผู้ป่วยโรคหัวใจและหลอดเลือด ความเชื่อที่ว่าเป็นผู้ป่วยที่เป็นผู้สูบบุหรี่จะปรับเป็น 0.5175 ซึ่งเพิ่มขึ้นจาก 0.199 ตามที่กำหนดไว้สำหรับความเป็นไปได้ในปัจจุบัน

อัตราส่วนระหว่างความน่าจะเป็นที่ปรับแล้วกับความน่าจะเป็นปัจจุบันแสดงให้เห็นถึงผลกระทบที่เกิดขึ้นจากเหตุการณ์ไม่พึงประสงค์ต่างๆ ในเหตุการณ์ด้านบนเรียกว่า Normalized Likelihood (NL) ตามที่อธิบายไว้ในส่วนที่ 2.1.5 แสดงดังสมการ (2-5) ถ้ามีการกำหนดตัวแปรมากกว่าหนึ่งสาเหตุ NL จะใช้เพื่อเปรียบเทียบเพื่อระบุความเสี่ยงหลักที่ทำให้เกิดโรคที่สนใจซึ่งเป็นตัวแปรเป้าหมายในแต่ละโมเดล

3.3.3 การจัดลำดับปัจจัยเสี่ยง (Risk Prioritisation)

การใช้การวิเคราะห์แบบ Influence Analysis เป็นตัวจัดลำดับปัจจัยที่ส่งผลต่อการเกิดโรค งานวิจัยนี้เลือกใช้ Normalised Likelihood (Kjaerulff & Madsen, 2008) เพื่อใช้เป็นดัชนีในการเปรียบเทียบแต่ละปัจจัยที่ส่งผลต่อการเกิดโรค โดยพิจารณาจากความน่าจะเป็นที่เปลี่ยนไปของแต่ละปัจจัยเมื่อทราบว่าเป็นผู้ป่วย (Adjusted Probability) เทียบกับความน่าจะเป็นที่เกิดขึ้นในปัจจุบัน (Current Probability) แสดงดังการคำนวณ Normalised Likelihood (NL) ในสมการ (3-3)

$$NL = \frac{P(\varepsilon'|X)}{P(\varepsilon')} \qquad \textbf{(3-3)}$$

เมื่อ

$P(\varepsilon'|X)$ คือ ความน่าจะเป็นที่เปลี่ยนไปของแต่ละปัจจัยเมื่อทราบว่าเป็นผู้ป่วย

$P(\varepsilon')$ คือ เทียบกับความน่าจะเป็นที่เกิดขึ้นโดยทั่วไป (Marginal Prior)

3.3.4 การจัดกลุ่มปัจจัยเสี่ยง (Risk Classification)

จากการแบ่งกลุ่มปัจจัยเสี่ยงในงานวิจัยทั่วไป มักใช้ในรูปแบบตารางประเมินความเสี่ยง (Risk Matrix) ที่ใช้องค์ประกอบของความเสี่ยงซึ่งประกอบด้วยความน่าจะเป็น (Probability) และผลกระทบของแต่ละปัจจัย (Impact) ในการพิจารณาแบ่งกลุ่มปัจจัยเสี่ยงเพื่อนำสู่การพิจารณากำหนดนโยบายเพื่อจัดการปัจจัยเสี่ยงเหล่านั้น โดยกำหนดตารางประเมินความเสี่ยง (Risk Matrix) ดังรูป 3-4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Impact/consequence

High	Transfer	Avoidance
Low	Retain/ Absorption	Reduction/ Mitigation
	Low	High

Perceived probability

รูป 3-4 การจับคู่การจัดการความเสี่ยงกับมิติของความน่าจะเป็นและผลกระทบ

ในการศึกษาครั้งนี้จะนำทั้งความน่าจะเป็น (Probability) ของการเกิดแต่ละปัจจัยที่นำเสนอด้วยค่าความน่าจะเป็นของแต่ละตัวแปรที่จะเกิดขึ้น (Marginal Probability) ของสถานะ (State) ที่สนใจที่อาจส่งผลต่อการเกิดโรคเบาหวาน และผลกระทบของแต่ละปัจจัย (Impact) จากค่า Normalised Likelihood (NL) ซึ่งคำนวณจาก เปอร์เซ็นต์ของเสียที่ตัวแปรที่สนใจเกิดความน่าจะเป็นเพิ่มขึ้นเมื่อทราบว่าเป็นผู้ป่วยโรคเบาหวานเทียบกับความน่าจะเป็นปัจจุบันของแต่ละตัวแปร ค่าความน่าจะเป็นดังกล่าวทั้ง 2 ส่วนถูกนำมาเป็นตัวแทนของความน่าจะเป็นและผลกระทบตามองค์ประกอบของความเสี่ยง (Risk) แสดงได้ดัง รูป 3-4

% of increase (Normalised Likelihood : NL)

High	Transfer	Avoidance
Low	Retain/ Absorption	Reduction/ Mitigation
	Low	High

ความน่าจะเป็นของแต่ละตัวแปรที่จะเกิดขึ้น (Marginal probability)

รูป 3-5 ตารางประเมินความเสี่ยง (Risk Matrix) สำหรับงานวิจัยนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลจากการแบ่งกลุ่มปัจจัยเสี่ยงตาม ขตารางประเมินความเสี่ยง (Risk Matrix) จะช่วยให้สามารถแบ่งกลุ่มปัจจัยเสี่ยงเป็น 4 กลุ่ม โดยนำสู่การกำหนดใช้กลยุทธ์การลดความเสี่ยงที่เกิดขึ้นจากปัจจัยต่างๆ ได้อย่างเหมาะสม โดยกำหนดแผนกลยุทธ์ในการจัดการความเสี่ยง ไว้ 4 กลุ่มกลยุทธ์ (Vose, 2008) ดังนี้ คือ

1. การหลีกเลี่ยงความเสี่ยง (Risk Avoidance) เหมาะสมกับความเสี่ยงที่อาจเกิดขึ้นได้บ่อยครั้งและเมื่อเกิดขึ้นอาจส่งผลกระทบสูงกับการเกิดโรคที่สนใจในตัวแปรเป้าหมาย

2. การถ่ายโอนความเสี่ยง (Risk Transfer) เหมาะสำหรับความเสี่ยงที่พบไม่บ่อยนักแต่ผลกระทบสูง จึงมีวัตถุประสงค์เพื่อเปลี่ยนผลกระทบให้กับผู้อื่นโดยการซื้อประกันหรือการจ้างบุคคลภายนอกมาดำเนินการแทน (Outsource)

3. การลดความเสี่ยง (Risk Reduction/ Mitigation) เหมาะสมกับความเสี่ยงที่พบบ่อยแต่มีผลกระทบไม่รุนแรง (ความน่าจะเป็นสูงและผลกระทบต่ำ) โดยพิจารณาการจัดการโดยต้องคำนึงถึงผลประโยชน์ที่จะได้รับกับต้นทุนและค่าใช้จ่าย

4. การคงความเสี่ยง (Risk Retention) เหมาะสมกับความเสี่ยงที่ไม่สำคัญเพราะมีโอกาสและผลกระทบน้อยมากเมื่อเทียบกับต้นทุนในการควบคุม

3.3.5 การพยากรณ์การเกิดโรค

จากโมเดลเครือข่ายแบบเบย์จะนำมาทดสอบความสามารถของโมเดลในการพยากรณ์การเกิดโรค เพื่อดูความสามารถในการทำนายกลุ่มของผู้ที่เป็นและไม่เป็นโรคเรื้อรังที่ไม่ติดต่อ โดยผลจากโมเดลจะให้ค่าพยากรณ์ในรูปความน่าจะเป็นของการเกิดโรคเพื่อให้นำมาจักกลุ่มของการเป็นโรคที่สนใจหรือไม่ หากการจัดกลุ่มของโมเดลให้ค่ากลุ่มที่ถูกต้องมากก็เป็นการบ่งชี้ว่าโมเดลดังกล่าวมีความสามารถในการพยากรณ์หรือการจำแนกค่อนข้างสูง สำหรับงานวิจัยนี้ผู้วิจัยจะนำเสนอผลการวิเคราะห์ทั้งในรูป Confusion matrix ค่า F1 และ เส้นโค้ง ROC และ ค่า AUC เพื่อใช้เปรียบเทียบความสามารถในการพยากรณ์ของโมเดลเครือข่ายแบบเบย์ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4 ผลการวิจัย

ในบทนี้จะนำเสนอผลจากการวิจัยที่ได้จากการสร้างโมเดลเครือข่ายแบบเบย์เพื่อนำไปสู่การวิเคราะห์และนำเสนอผลการวิเคราะห์ตามวัตถุประสงค์ที่กำหนดไว้สำหรับงานวิจัยนี้ ผลจากการวิเคราะห์ดังกล่าวอาจนำไปสู่การกำหนดแนวทางการสนับสนุนการวิเคราะห์ความเสี่ยงของการเกิดโรคไม่ติดต่อที่สำคัญอื่นๆ ต่อไปในอนาคต

1. วิเคราะห์ปัจจัยเสี่ยงหรือตัวบ่งชี้ความเสี่ยงที่สำคัญต่อการเกิดโรคไม่ติดต่อที่สำคัญของคนไทยจากข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกายจากโมเดลเครือข่ายแบบเบย์ (Bayesian Network Model)
    a. จัดลำดับปัจจัยเสี่ยงที่อาจก่อให้เกิดโรคไม่ติดต่อที่สำคัญแต่ละโรคด้วยการอนุมานแบบ Diagnostic Inference
    b. การจัดกลุ่มความเสี่ยงเพื่อกำหนดนโยบายที่เหมาะสมของความเสี่ยงแต่ละกลุ่ม
2. พยากรณ์การเกิดโรคไม่ติดต่อที่สำคัญของคนไทยจากข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกายจากโมเดลเครือข่ายแบบเบย์ (Bayesian Network)

4.1 โรคเบาหวาน

4.1.1 ตัวแปรและนิยาม

ตัวแปรที่ใช้ในการสร้างโมเดลเครือข่ายแบบเบย์เพื่อศึกษาการเกิดโรคเบาหวานกำหนดขึ้นโดยการสัมภาษณ์ แพทย์ที่เข้าร่วมในโครงการวิจัยในฐานะผู้เชี่ยวชาญ (Experts) และแพทย์ผู้ดำเนินการโครงการการเก็บข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย ครั้งที่ 4 (สำนักงานสำรวจสุขภาพประชาชนไทย, 2552) กำหนดตัวแปรสาเหตุการเกิดโรคเบาหวานที่เกี่ยวข้องทั้งสิ้น 7 ตัวแปร และตัวแปรที่ระบุการเกิดโรคเบาหวานที่ได้มาจากผลการตรวจทางห้องปฏิบัติการ ดังนั้นจึงนำตัวแปรทั้ง 8 ตัวแปร มาทำการศึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 4-1 ตัวแปร นิยาม และกลุ่มระดับตัวแปร (State)

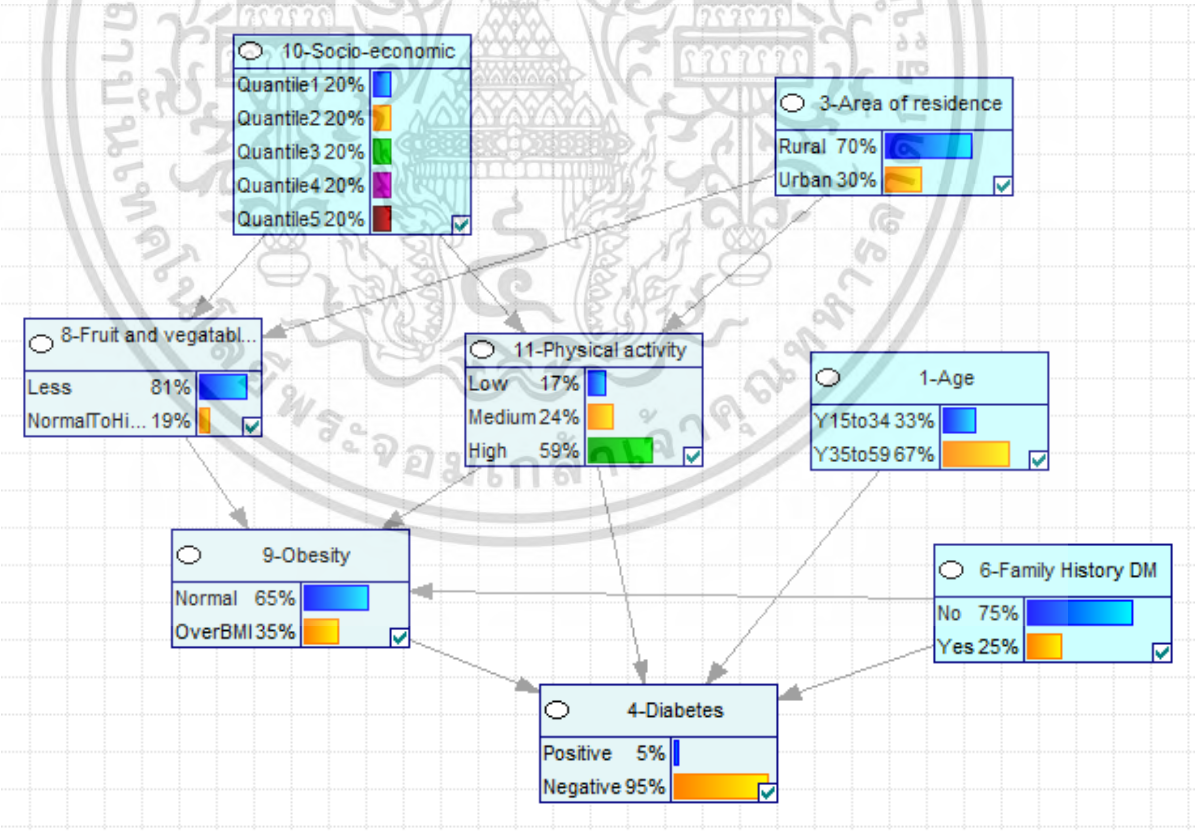
ลำดับ	ตัวแปรและระดับ	นิยาม
1	**Age**	กลุ่มอายุ
	15-34	อายุระหว่าง 15-34 ปี
	35-59	อายุระหว่าง 35-59 ปี
2	**Area of residence**	เขตที่พักอาศัย
	Urban	ในเขตเทศบาล
	Rural	นอกเขตเทศบาล
3	**Diabetes**	ภาวะการเป็นโรคเบาหวาน
	Positive	การตรวจเลือดหลังอดอาหารนาน 12 ชั่วโมง (Fasting Plasma Glucose, FPG) ≥ 126 mg/dl หรือ เคยได้รับการบอกว่าเป็นโรคเบาหวาน
	Negative	การตรวจเลือดหลังอดอาหารนาน 12 ชั่วโมง (Fasting Plasma Glucose, FPG) < 126 mg/dl
4	**Family History DM**	ประวัติการเป็นโรคเบาหวานของครอบครัว (สายตรง)
	Yes	มีญาติสายตรงได้แก่ พ่อ แม่ หรือพี่น้องมีประวัติเป็นโรคเบาหวาน
	No	ไม่มีญาติสายตรงได้แก่ พ่อ แม่ หรือพี่น้องมีประวัติเป็นโรคเบาหวาน
5	**Fruit and Vegetable consumption หรือ Veg**	การบริโภคผักและผลไม้
	Less	ผักผลไม้ไม่เพียงพอ < 5 servings/day
	Normal to high	> 5 servings/day
6	**Obesity**	ภาวะโรคอ้วน
	Normal	ทั้ง BMI และรอบเอวต่ำกว่าเกณฑ์
	Over BMI	BMI ≥ 25 (รอบเอวเกินเกณฑ์)
7	**Socio-economic**	ความมั่นคงทางเศรษฐกิจและสังคม
	Quintile1	ตัวแปรดัชนี Wealth index score คำนวณ อยู่ใน Quintile 1
	Quintile2	ตัวแปรดัชนี Wealth index score คำนวณ อยู่ใน Quintile 2
	Quintile3	ตัวแปรดัชนี Wealth index score คำนวณ อยู่ใน Quintile 3
	Quintile4	ตัวแปรดัชนี Wealth index score คำนวณ อยู่ใน Quintile 4
	Quintile5	ตัวแปรดัชนี Wealth index score คำนวณ อยู่ใน Quintile 5
8	**Physical activity**	การเคลื่อนไหวร่างกายที่มีการใช้พลังงานในร่างกายโดยคำนึงถึงระดับความหนักเบาของกิจกรรมทางกาย
	Low	ระดับของการมีกิจกรรมทางกายต่ำกว่าเกณฑ์ระดับปานกลางและระดับมาก
	Medium	a. มีกิจกรรมทางกายไม่มากถึงระดับมาก และ a. มีกิจกรรมอย่างหนัก ≥ 3 วัน/สัปดาห์ และเวลา ≥ 20 นาทีต่อวัน b. มีกิจกรรมปานกลางหรือเดิน รวม ≥ 5 วัน อย่างน้อยวันละ 30 นาทีต่อวัน และ total MET-นาทีต่อสัปดาห์ ≥ 1,500 หรือ c. มีกิจกรรมทางกายอย่างหนักและปานกลาง หรือ เดิน รวม ≥ 5 วัน/สัปดาห์ และ total MET-นาทีต่อสัปดาห์ ≥ 600
	High	มีกิจกรรมทางกายอย่างหนัก ≥ 3 วัน/สัปดาห์ และ total MET-นาทีต่อสัปดาห์ ≥ 1,500 หรือ มีกิจกรรมทางกายอย่างหนักหรือปานกลางรวม ≥ 7 วัน/สัปดาห์ และ total MET-นาทีต่อสัปดาห์ ≥ 3,000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เริ่มต้นจากการกำหนดตัวแปรเป้าหมาย (Focal Variable) ซึ่งก็คือ การเกิดโรคเบาหวาน (Diabetes) โดยประกอบด้วยตัวแปรที่เป็นสาเหตุของการเกิดโรคเบาหวานทั้งทางตรงและทางอ้อม ทั้งสิ้น 7 ตัวแปร โดยกำหนดนิยามของแต่ละตัวแปรจากรายงานการสำรวจสุขภาพของประชาชนไทย โดยการตรวจร่างกาย ครั้งที่ 4 แสดงดังตาราง 4-1

4.1.2 โครงสร้างโมเดลเครือข่ายแบบเบย์ของการเกิดโรคเบาหวาน

การสร้างโมเดลเครือข่ายแบบเบย์จากความคิดเห็นแพทย์ที่เข้าร่วมในโครงการวิจัยในฐานะผู้เชี่ยวชาญ (Experts) และแพทย์ผู้ดำเนินการโครงการเก็บข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย ครั้งที่ 4 ร่วมกำหนดโครงสร้างในเชิงเหตุและผลเพื่อแสดงสาเหตุของการเกิดโรคเบาหวานทั้งในรูปสาเหตุทางตรงและทางอ้อม โมเดลผ่านการตรวจสอบและปรับปรุงหลายครั้งจนได้รูปแบบโครงสร้างของโมเดลเครือข่ายแบบเบย์ที่เหมาะสม นอกจากนี้ยังมีข้อมูลจากการสำรวจฯ สนับสนุนการคำนวณเพื่อใช้กำหนดค่าความน่าจะเป็น จนกระทั่งได้โครงสร้างโมเดลแสดงได้ดังต่อไปนี้



รูป 4-1 โครงสร้างโมเดลเครือข่ายแบบเบย์สำหรับการเกิดโรคเบาหวาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การโมเดลที่สร้างขึ้นจากผู้เชี่ยวชาญพบว่าปัจจัยที่อาจส่งผลต่อการเกิดโรคเบาหวานโดยตรงมากจากปัจจัยด้านอายุ การมีภาวะโรคอ้วน การเคลื่อนไหวร่างกายที่มีการใช้พลังงานในร่างกายโดยคำนึงถึงระดับความหนักเบาของกิจกรรมทางกาย และประวัติการเป็นโรคเบาหวานของครอบครัว (สายตรง) ในขณะที่ความมั่นคงทางเศรษฐกิจและสังคม เขตที่พักอาศัย อาจส่งผลต่อพฤติกรรมการบริโภคผักและผลไม้ และการเคลื่อนไหวร่างกายที่มีการใช้พลังงานในร่างกายโดยคำนึงถึงระดับความหนักเบาของกิจกรรมทางกาย ที่ทั้ง 2 ปัจจัยอาจส่งผลสู่ภาวะมีภาวะโรคอ้วนที่อาจส่งผลต่อการเกิดโรคเบาหวาน

4.1.3 การวิเคราะห์ปัจจัยเสี่ยงหรือตัวบ่งชี้ความเสี่ยงที่สำคัญต่อการเกิดโรคเบาหวาน

การวิเคราะห์เพื่อระบุหาตัวบ่งชี้ที่สำคัญของการเกิดโรคเบาหวานจากโมเดลเครือข่ายแบบเบย์ (Bayesian Network) เป็นการวิเคราะห์เพื่อกำหนดแผนการดำเนินงานเพื่อจัดการกับต้นเหตุของปัญหาการเกิดโรคเบาหวานได้อย่างเหมาะสม เพื่อให้สามารถกำหนดแนวทางการใช้ทรัพยากรที่มีอยู่อย่างเหมาะสม (Stephan & Badr, 2007) ซึ่งเป็นแนวทางที่สำคัญในการจัดการความเสี่ยง ดังนั้นสำหรับการวิเคราะห์ในการศึกษาครั้งนี้จะนำเสนอแนวทางการวิเคราะห์ความเสี่ยงการเกิดโรคเบาหวานจากโมเดลเครือข่ายแบบเบย์ที่สร้างขึ้นจากข้อมูลภาพรวมของทั้งประเทศเพื่อจัดเรียงลำดับปัจจัยเสี่ยงการเกิดโรคเบาหวานในระดับประชากร เพื่อช่วยในการกำหนดนโยบายในการแก้ปัญหาที่เหมาะสมสำหรับประเทศไทย ภายใต้งบประมาณที่มีอยู่อย่างจำกัดเพื่อให้เกิดประสิทธิภาพสูงสุด

ตาราง 4-2 สถานะของแต่ละตัวแปรที่ใช้ในการวิเคราะห์โรคเบาหวาน

ตัวแปร	ชื่อตัวแปร (Variable Name)	สถานะที่สนใจ (State)
กลุ่มอายุ	Age	35-59
ประวัติการเป็นโรคเบาหวานของครอบครัว (สายตรง)	Family History DM	Yes
เขตที่พักอาศัย	Area of residence	Urban
ความมั่นคงทางเศรษฐกิจและสังคม	Socio-economic	Quintile 4
การบริโภคผักและผลไม้	Fruit and Vegetable consumption	Normal to high
ภาวะโรคอ้วน	Obesity	Over BMI
การเคลื่อนไหวร่างกาย	Physical activity	Medium

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการวิเคราะห์เริ่มต้นด้วยการเลือกพิจารณาสถานะ (State) ที่สนใจของแต่ละตัวแปรปัจจัยสาเหตุ โดยจากการพิจารณาสถานะ (State) ที่ส่งผลต่อการเกิดโรคเบาหวานในแต่ละตัวแปร โดยกำหนดได้ดังตาราง 4-2

งานวิจัยนี้นำเสนอการจัดลำดับปัจจัยเสี่ยง (Risk Prioritisation) ต่อการเกิดโรคเบาหวานของคนไทย โดยจะเรียงลำดับปัจจัยที่เป็นปัจจัยเสี่ยงต่อการเกิดโรคเบาหวานมากที่สุดไปน้อยสุด ซึ่งจะนำเสนอในหัวข้อ 4.1.3.1 นอกจากนี้ยังมีการจัดกลุ่มปัจจัยเสี่ยง (Risk Classification) ที่อาจส่งผลต่อการเกิดโรคเบาหวานจากทฤษฎีเรื่องวิเคราะห์ความเสี่ยง โดยอาศัยทฤษฎีตารางประเมินความเสี่ยง (Risk Matrix) เพื่อนำผลจากการวิเคราะห์ช่วยให้หน่วยงานที่รณรงค์เกี่ยวกับการป้องกันโรคไม่ติดต่อมีทิศทางในการทำงานด้านการวางแผนนโยบายที่สามารถนำมาใช้กับกลุ่มประชากรไทยในอนาคต แสดงดังหัวข้อ 4.1.3.2 สำหรับในหัวข้อสุดท้ายจะนำเสนอการใช้โมเดลเครือข่ายแบบเบย์ในการพยากรณ์การเกิดโรคเบาหวาน แสดงในหัวข้อ 4.1.3.4

4.1.3.1 การจัดลำดับปัจจัยเสี่ยง (Risk Prioritisation) ต่อการเกิดโรคเบาหวานของคนไทย

การวิเคราะห์จากโมเดลเครือข่ายแบบเบย์ (Bayesian Network) ของการเกิดโรคเบาหวาน เพื่อหาคำตอบของคำถามที่ว่า “อะไรเป็นปัจจัยหลักที่ส่งผลต่อการเป็นโรคเบาหวาน?”

คำนวณ Normalised Likelihood ซึ่งเป็นอัตราส่วนระหว่างความน่าจะเป็นที่เปลี่ยนไปของแต่ละปัจจัยเมื่อทราบว่าเป็นผู้ป่วยโรคเบาหวาน (Adjusted Probability) และความน่าจะเป็นที่เกิดขึ้นในปัจจุบัน (Current Probability) ในสมการ (2-7) ดังนั้นเราจะคำนวณค่า Normalised Likelihood และ เปอร์เซ็นต์ที่เพิ่มขึ้น และกำหนดค่าดังกล่าวในการเรียงลำดับปัจจัยเสี่ยงที่เกี่ยวข้องต่อการเกิดโรคเบาหวานแสดงได้ดังตาราง 4-3

ตัวอย่างเช่น โอกาสที่จะพบผู้มีประวัติการเป็นโรคเบาหวานของครอบครัวสายตรง (Family History DM) เพิ่มขึ้นจาก 0.2524 เป็น 0.5220 เมื่อทราบว่าเขาเป็นโรคเบาหวาน ทำให้มีอัตราส่วนที่เพิ่มขึ้น เป็น 2.0682 เท่า คิดเป็นโอกาสที่เพิ่มขึ้น 106.8219%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 4-3 ผลการเรียงลำดับปัจจัยเสี่ยงต่อการเกิดโรคเบาหวานของประชาชนไทย

ชื่อตัวแปร (Variable Name)	สถานะที่สนใจ (State)	ความน่าจะเป็น		Normalised likelihood	% of Increase	ลำดับ (Rank)
		ปัจจุบัน (Current)	เปลี่ยนไป (Adjusted)			
Age	35-59 year	0.6664	0.8942	1.3419	34.1851	3
Family History DM	Yes	0.2524	0.5220	2.0682	106.8219	1
Area of residence	Urban	0.3016	0.3025	1.0028	0.2824	7
Socio-economic	Quintile 4	0.2000	0.2007	1.0034	0.3414	6
Fruit and Vegetable consumption	Normal to high	0.8107	0.8143	1.0044	0.4407	5
Obesity	Over BMI	0.3547	0.5941	1.6749	67.4931	2
Physical activity	Medium	0.1651	0.1743	1.0556	5.5588	4
Diabetes	Positive	0.0528				

จากตารางข้างต้นพบว่าปัจจัยด้านประวัติการเป็นโรคเบาหวานของครอบครัวสายตรง (Family History DM) เป็นปัจจัยที่มีอัตราส่วนการเปลี่ยนแปลงความน่าจะเป็นเพิ่มขึ้นเมื่อทราบว่าเป็นโรคเบาหวานเทียบกับความน่าจะเป็นที่เกิดขึ้นทั่วไปคิดเป็นร้อยละ 106.8219 รองลงมาคือการเป็นโรคอ้วน (Obesity) มีความน่าจะเป็นเพิ่มขึ้นคิดเป็นร้อยละ 37.493 ปัจจัยเสี่ยงอันดับ 3 ได้แก่ ปัจจัยด้านอายุ (Age) ในกลุ่มอายุ 35-59 ปี มีความน่าจะเป็นเพิ่มขึ้นคิดเป็นร้อยละ 34.1851 ปัจจัยเสี่ยงอันดับ 4 ได้แก่ปัจจัยด้านการเคลื่อนไหวทางกาย (Physical Activity) ในระดับปานกลาง มีความน่าจะเป็นเพิ่มขึ้นคิดเป็นร้อยละ 5.5588 ปัจจัยเสี่ยงอันดับ 5 ได้แก่ การบริโภคผักและผลไม้ (Fruit And Vegetable Consumption) ในระดับ น้อย มีความน่าจะเป็นเพิ่มขึ้นคิดเป็นร้อยละ 0.4407 ปัจจัยเสี่ยงอันดับ 6 ได้แก่สถานะทางสังคม (Socio-Economic) ในระดับ Quintile 4 มีความน่าจะเป็นเพิ่มขึ้นคิดเป็นร้อยละ 0.3414 และปัจจัยเสี่ยงอันดับสุดท้ายคือด้านที่อยู่อาศัย (Area Of Residence) ในเขตเทศบาล มีความน่าจะเป็นเพิ่มขึ้นคิดเป็นร้อยละ 0.2824

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.3.2 การจัดกลุ่มปัจจัยเสี่ยง (Risk classification) ที่อาจส่งผลต่อการเกิดโรคเบาหวานของคนไทย

การวิเคราะห์จากโมเดลเครือข่ายแบบเบย์ (Bayesian Network) ของการเกิดโรคเบาหวาน เพื่อหาคำตอบของคำถามที่ว่า “ปัจจัยความเสี่ยงการเกิดโรคเบาหวานของคนไทยควรจัดการอย่างไร”

การกำหนดแนวทางการจัดการสาเหตุการเกิดโรคเบาหวานใช้แนวทางการพิจารณาจากการกำหนดตารางประเมินความเสี่ยง (Risk Matrix) โดยแบ่งกลุ่มสาเหตุความเสี่ยงของการเกิดโรคจาก 2 องค์ประกอบหลักได้แก่ ความน่าจะเป็นของการเกิดความเสี่ยงของแต่ละสาเหตุการเกิดโรคซึ่งในที่นี้เรียกปัจจัยเสี่ยงซึ่งนำเสนอในรูปความน่าจะเป็นที่เกิดขึ้นในปัจจุบัน (Current Probability) โดยทั่วไป และผลกระทบที่เกิดจากความน่าจะเป็นที่เปลี่ยนไปเมื่อทราบว่าเป็นผู้ป่วยโรคเบาหวานที่นำเสนอในรูปของ % ของความน่าจะเป็นที่เพิ่มขึ้นของแต่ละปัจจัยเสี่ยงเมื่อทราบว่าเป็นผู้ป่วยเบาหวาน (% Increase) ทั้งสองปัจจัยเป็นตัวกำหนดแกน X และ Y

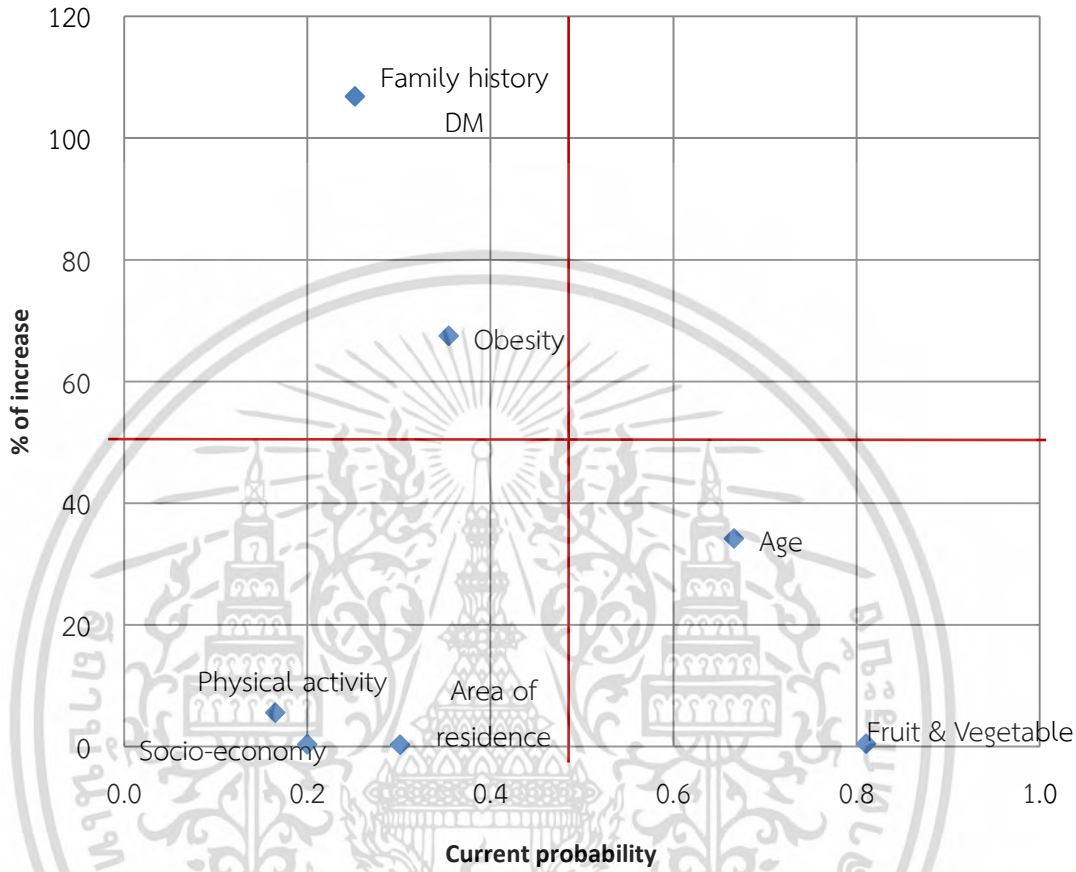
พื้นที่ในกราฟสามารถแบ่งได้เป็น 4 ควอดรันต์ (Quadrant) หรือ จตุภาค ตามการแบ่งค่าสูงและต่ำของทั้ง 2 ปัจจัยดังกล่าว จากนั้นเลือกแต่ละปัจจัยที่มีสถานะ (State) ที่ให้ค่า Normalised Likelihood มากกว่า 1 เป็นตัวแทนแต่ละปัจจัยเพื่อนำมาแบ่งกลุ่ม (ดังที่ได้แสดงดังตาราง 4-2) แสดงผลการแบ่งกลุ่มปัจจัยเสี่ยงได้ดังรูป 4-2 จากรูปจะเห็นได้ว่าหากกำหนดจุดตัดเพื่อแบ่งค่าสูงและต่ำที่ 50% ซึ่งประกอบด้วย กลุ่ม High Probability & High Impact กลุ่ม Low Probability & High Impact กลุ่ม High Probability & Low Impact กลุ่ม Low Probability & Low Impact โดยพบว่าไม่มีปัจจัยใดอยู่ในกลุ่ม High Probability & High Impact ส่วนในกลุ่มที่เหลือมีผลได้ดังนี้

ส่วนปัจจัยด้านประวัติการเป็นโรคเบาหวานของครอบครัวสายตรง (Family History DM) และปัจจัยด้านการเป็นโรคอ้วน (Obesity) ถึงแม้มีพบในกลุ่มประชากรไทยโดยทั่วไปมีโอกาสพบกลุ่มเหล่านี้ไม่มาก และหากทราบว่าคนๆ นั้นเป็นโรคเบาหวานจะทำให้โอกาสที่จะพบว่ามีประวัติการเป็นโรคเบาหวานของครอบครัวสายตรง (Family History DM) และปัจจัยด้านการเป็นโรคอ้วน (Obesity) สูงขึ้นมากเมื่อเทียบกับกลุ่มประชากรทั่วไป ด้วยเหตุผลนี้จึงกำหนดทั้ง 2 ตัวแปรเป็นตัวแปรที่ส่งผลต่อการเกิดโรคเบาหวานมากกว่าตัวแปรอื่น จัดอยู่ในกลุ่ม (Low Probability & High Impact) การจัดการลดความเสี่ยงให้กลับบุคคลกลุ่มนี้ทำได้โดยการ Transfer ความเสี่ยง เช่น แนะนำให้มีการทำประกันสุขภาพที่สามารถครอบคลุมการรักษาโรคเบาหวาน

โอกาสที่จะพบกลุ่มประชากรที่มีอายุ (Age) ในช่วง 35-59 ปี (ไม่รวมกลุ่มผู้สูงอายุ) พบว่าเป็นในกลุ่มประชากรส่วนใหญ่ (เนื่องจากสังคมไทยเข้าสู่สังคมสูงอายุ) และโอกาสที่จะพบผู้บริโภคผักและผลไม้ในปริมาณน้อย มีในระดับสูง และหากทราบว่าคนๆ นั้นเป็นโรคเบาหวานจะทำให้โอกาสที่จะพบว่าเป็นคนในกลุ่มอายุในช่วง 35-59 ปี และปัจจัยด้านการบริโภคผักและผลไม้ในปริมาณน้อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพิ่มขึ้นไม่สูงมากนักเมื่อเทียบกับกลุ่มประชากรทั่วไป (High Probability & Low Impact) การจัดการความเสี่ยงจะเลือกวิธีการ Reduction/Mitigation เช่น การส่งเสริมให้บริโภคผักและผลไม้มากขึ้น



รูป 4-2 ผลการจัดกลุ่มความเสี่ยงการเกิดโรคเบาหวาน

โอกาสที่จะพบกลุ่มประชากรที่อาศัย (Area Of Residence) อยู่ในเมือง หรือมีสถานะทางเศรษฐกิจและสังคม (Socio-Economic) ในระดับ 4 หรือมีกิจกรรมทางกาย (Physical Activity) ในระดับต่ำ มีความน่าจะเป็นไม่สูงนัก และหากทราบว่าคนๆ นั้นเป็นโรคเบาหวานจะทำให้โอกาสที่จะพบว่าเป็นคนในกลุ่มที่อาศัย (Area Of Residence) อยู่ในเมือง หรือมีสถานะทางเศรษฐกิจและสังคม (Socio-Economic) ในระดับ 4 หรือมีกิจกรรมทางกาย (Physical Activity) ในระดับต่ำ สูงขึ้นเพียงเล็กน้อยเมื่อเทียบกับกลุ่มประชากรทั่วไป (Low Probability & Low Impact) การจัดการความเสี่ยงจะเลือกวิธีการ Retain/Absorption โดยอาจไม่ต้องให้ความสำคัญกับปัจจัยเหล่านี้ และอาจยังคงอยู่ดังเดิม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.4 พยากรณ์การเกิดโรคเบาหวานของคนไทย

การสร้างตัวแบบเพื่อใช้พยากรณ์การเกิดโรคเบาหวานด้วยโมเดลเครือข่ายแบบเบย์ดำเนินการโดยใช้ตัวแบบที่สร้างขึ้นกับข้อมูลที่ใช้ในการสร้างตัวแบบ (Training Dataset) และข้อมูลที่ใช้สำหรับการทดสอบ (Testing Dataset) โดยใช้อัตราส่วน 70:30 นอกจากนี้ยังแสดงผลการพยากรณ์จากชุดข้อมูลทั้งหมด ดังนั้นผลการวิเคราะห์ที่นำเสนอในหัวข้อนี้จึงเป็นผลที่ได้จากชุดข้อมูลจากการทดสอบ (Testing Dataset) เป็นหลัก เพื่อให้เชื่อมั่นได้ว่าโมเดลมีความสามารถในการจำแนกกลุ่มการพยากรณ์การเกิดโรคเบาหวานหรือไม่ (Positive Or Negative) ได้ถูกต้องมากน้อยเพียงใด แสดงผลที่ได้จาก คอนฟูชั่น แมททริก คะแนน F1 และค่า AUC จากเส้นโค้ง ROC

ผลจากตาราง 4-4 ทำให้ทราบว่าโมเดลเครือข่ายแบบเบย์มีความผิดพลาดในการจำแนกกลุ่มการเกิดโรคเบาหวานระหว่าง 5.68% – 19.50% โดยเปอร์เซ็นต์ความผิดพลาดในการจำแนกมีค่าเพิ่มขึ้นเมื่อกำหนดระดับความน่าจะเป็นที่ใช้ในการกำหนดการเกิดโรคเบาหวาน (Threshold) ที่มีค่าลดลง โดยพบว่า ระดับเกณฑ์ความน่าจะเป็นที่ใช้ในการกำหนดการเกิดโรคที่ Threshold = 0.5, 0.4, 0.3 จะให้ค่าความผิดพลาด (Error Rate) 5.68% เพิ่มขึ้นเป็น 6.59% ที่ระดับเกณฑ์ (Threshold) = 0.2 และเพิ่มขึ้นเป็น 19.50% ที่ระดับเกณฑ์ (Threshold) = 0.1 หรืออาจกล่าวได้ว่าค่าความถูกต้อง (Accuracy) ลดลงจาก 94.32% เป็น 80.50% เมื่อกำหนดระดับเกณฑ์ (Threshold) ลดลงจาก 0.5 เป็น 0.1

โมเดลเครือข่ายแบบเบย์ให้ค่าความจำเพาะ (Specificity) และค่าทำนายเมื่อผลเป็นลบ (NPV) ในระดับสูง เนื่องจากมีค่า True Negative ค่อนข้างสูง และยังพบว่า โมเดลมีค่าความจำเพาะลดลงเมื่อค่า ระดับเกณฑ์ (Threshold) ลดลง แต่จะให้ค่าอัตราความถูกต้องเชิงลบ (True Negative Rate) เพิ่มขึ้นเมื่อระดับเกณฑ์ (Threshold) ลดลง

โมเดลเครือข่ายแบบเบย์ให้ค่าความระลึก (Recall) และค่าความแม่นยำ (Precision) ต่ำ เนื่องจากมีค่า True Positive ค่อนข้างต่ำ แต่พบว่า ค่า True Positive จะมีค่ามากขึ้น แต่ True Negative จะมีค่าลดลง เมื่อลดระดับเกณฑ์ (Threshold) ลง ด้วยเหตุนี้จึงทำให้ค่าความระลึก (Recall) ลดลงแต่ค่าความแม่นยำ (Precision) เพิ่มขึ้นเมื่อลดระดับเกณฑ์ (Threshold) ลดลง เนื่องจากทั้ง 2 ค่า ดังกล่าวมีทิศทางความสัมพันธ์ที่ขัดแย้งกัน จึงใช้ค่า F1 เพื่อหาค่าเฉลี่ยของทั้ง 2 ค่า และพบว่าค่า F1 เพิ่มขึ้นจาก 0.094 เป็น 0.2449 เมื่อลดระดับเกณฑ์ (Threshold) ลง จาก 0.2 เป็น 0.1 แสดงว่าความสามารถในการพยากรณ์ของโมเดลดีขึ้นเมื่อลดเกณฑ์ (Threshold) ลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 4-4 คอนฟูชั่น แมททริก และค่าต่างๆ ที่ใช้ในการวัดความสามารถในการจำแนกโรคเบาหวาน ในระดับเกณฑ์ความน่าจะเป็นที่แตกต่างกัน

ค่าจริง (Actual)	ระดับเกณฑ์การจำแนก (Threshold Level)									
	0.5		0.4		0.3		0.2		0.1	
	ค่าพยากรณ์ (Predictive)									
	P	N	P	N	P	N	P	N	P	N
BNH model										
P	0	181	0	181	0	181	15	166	103	78
N	0	3,004	0	3,004	0	3,004	44	2,960	543	2,461
Error rate ความผิดพลาด	5.68%		5.68%		5.68%		6.59%		19.50%	
ความถูกต้อง (Accuracy)	0.9432		0.9432		0.9432		0.9341		0.8050	
ค่าความจำเพาะ Specificity	1.0000		1.0000		1.0000		0.9854		0.8192	
ค่าทำนายเมื่อผลเป็นลบ (NPV)	0.9432		0.9432		0.9432		0.9469		0.9693	
ความระลึก (Recall)	0.0000		0.0000		0.0000		0.0829		0.5691	
ค่าความแม่นยำ (Precision)	-		-		-		0.2542		0.1594	
F1 score	-		-		-		0.1250		0.2491	

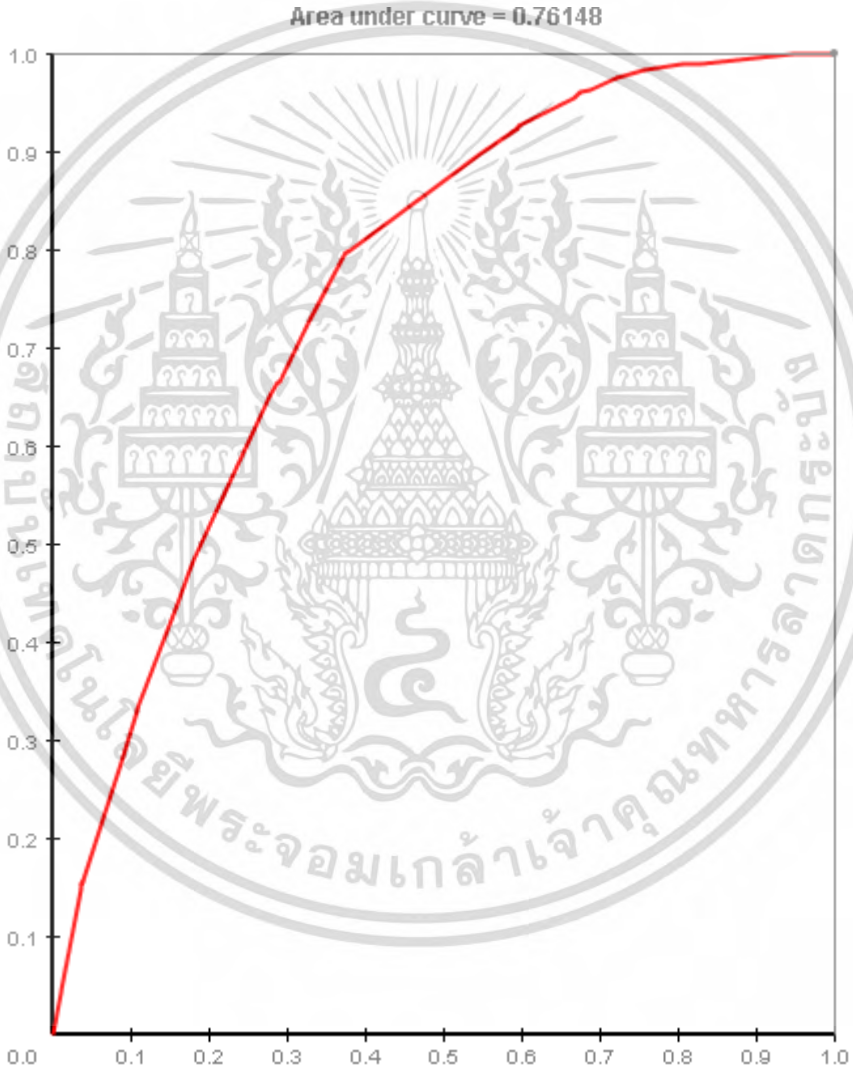
P: Positive; N: Negative

ตาราง 4-5 ค่า **AUC** จาก **Training Dataset, Testing Dataset** และ **Full Dataset** สำหรับโรคเบาหวาน

	Training Dataset	Testing Dataset	Full Dataset
AUC	0.76148	0.79659	0.7719

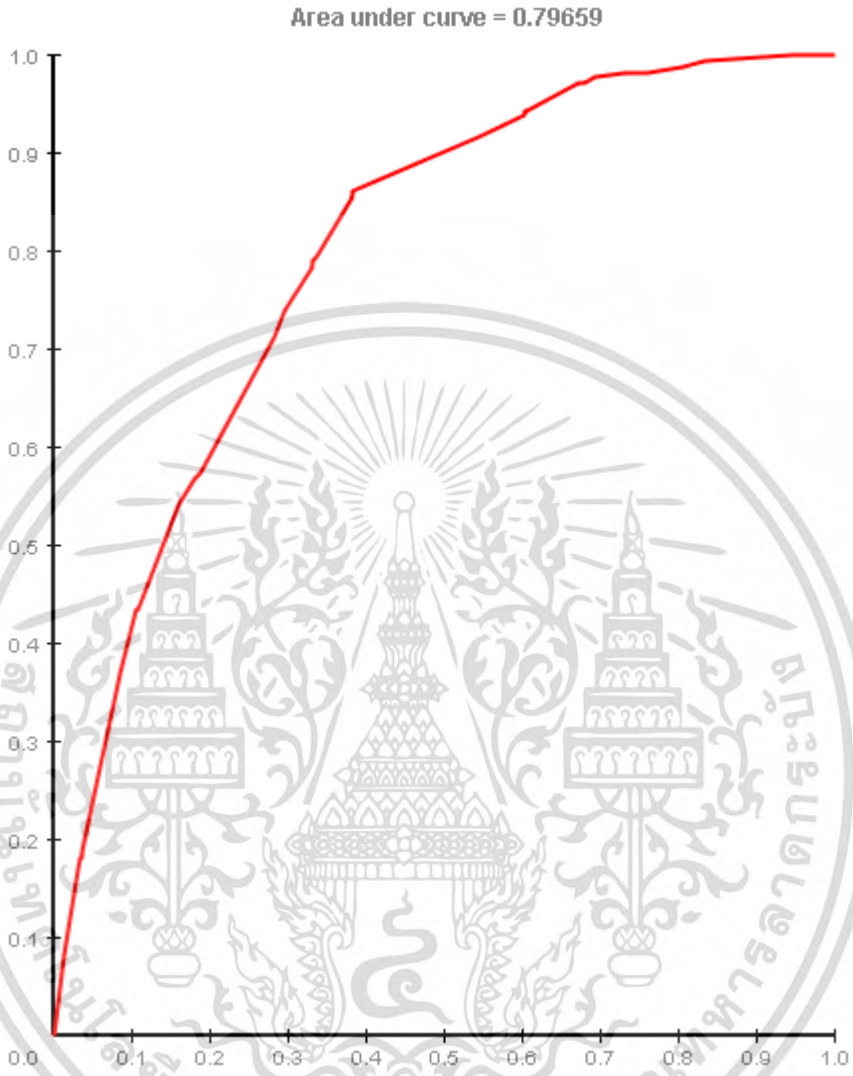
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตาราง 4-5 แสดงค่า AUC ซึ่งเป็นค่าที่แสดงพื้นที่ใต้กราฟของเส้นโค้ง ROC แสดงความสามารถในการทำนายผลอย่างถูกต้องเป็น 76.148 และ 79.659 ซึ่งคำนวณจาก Training และ Testing Dataset ตามลำดับ หากพิจารณาค่า AUC ของข้อมูลปกติที่ไม่ได้แบ่งเป็น Training และ Testing Dataset พบว่ามีค่าสามารถทายถูก 77.19 ซึ่งเป็นค่าที่อยู่ระหว่างค่าที่ได้จากชุดข้อมูล Training และ Testing Dataset



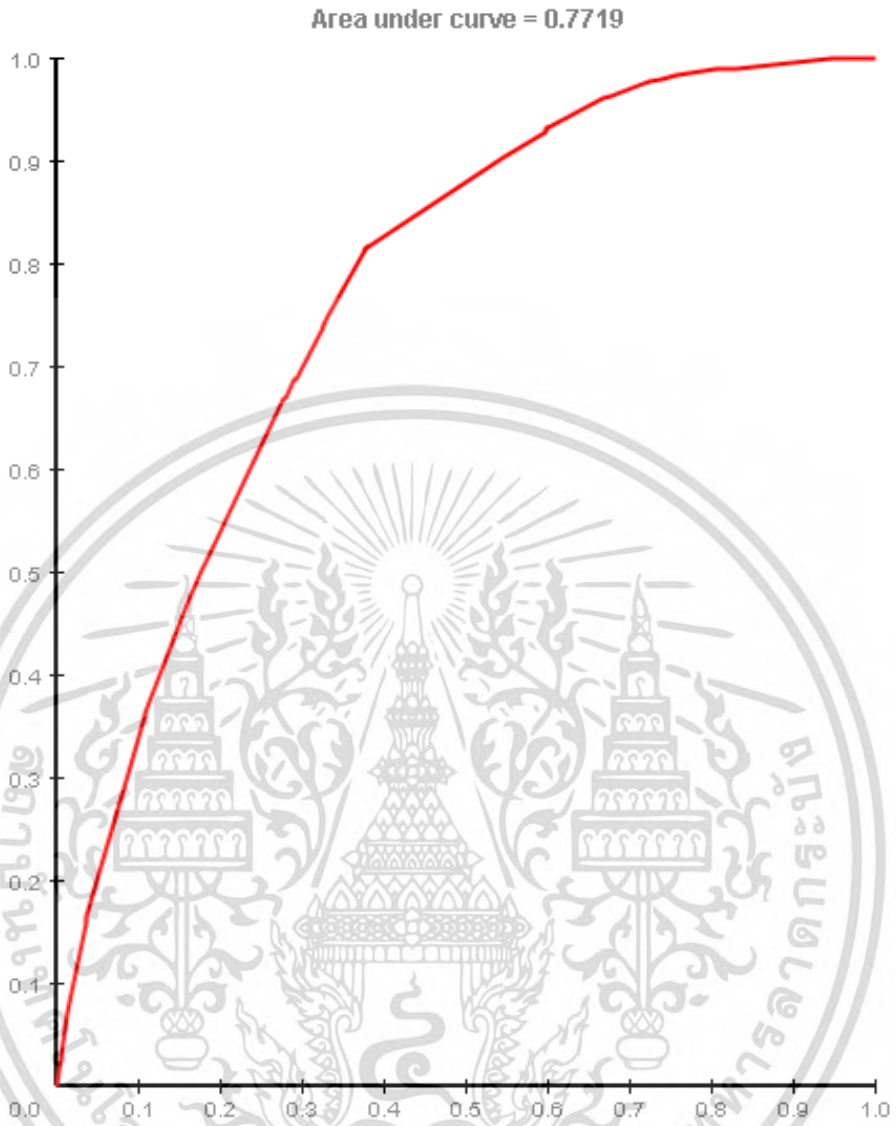
รูป 4-3 ROC Curve จาก Training Dataset สำหรับโรคเบาหวาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 4-4 ROC Curve จาก Testing Dataset สำหรับโรคเบาหวาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 4-5 ROC Curve จาก Full Dataset สำหรับโรคเบาหวาน

เมื่อพิจารณาเส้นโค้ง ROC ดังรูป 4-3 - 4-5 กราฟเป็นเส้นโค้ง ROC มีลักษณะโค้งเข้ามุมบนซ้ายมือ ที่ให้พื้นที่ใต้โค้งใกล้เคียงกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 โรคความดันโลหิตสูง

4.2.1 ตัวแปรและนิยาม

ตัวแปรที่ใช้ในการสร้างโมเดลเครือข่ายแบบเบย์เพื่อศึกษาการเกิดโรคความดันโลหิตสูง กำหนดขึ้นโดยการสัมภาษณ์แพทย์ที่เข้าร่วมในโครงการวิจัยในฐานะผู้เชี่ยวชาญ (Experts) และแพทย์ผู้ดำเนินการโครงการการเก็บข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย ครั้งที่ 4 (สำนักงานสำรวจสุขภาพประชาชนไทย, 2552, 2554) โดยกำหนดตัวแปรที่เป็นสาเหตุการเกิดโรคความดันโลหิตสูงทั้งสิ้น 10 ตัวแปร ดังนั้นข้อมูลจากตัวแปรทั้ง 10 ตัว จะนำมากำหนดโครงสร้างของโมเดลเครือข่ายแบบเบย์

เริ่มต้นจากการกำหนดตัวแปรเป้าหมาย (Focal Variable) ซึ่งก็คือ การเกิดโรคความดันโลหิตสูง (Hypertension) และตัวแปรที่เป็นสาเหตุของการเกิดโรคความดันโลหิตสูงทั้งทางตรงและทางอ้อมทั้งสิ้น 9 ตัวแปร โดยกำหนดนิยามของแต่ละตัวแปรโดยมีการดัดแปลงนิยามของระดับตัวแปรแต่ละตัวจากรายงานการสำรวจสุขภาพของประชาชนไทยโดยการตรวจร่างกาย ครั้งที่ 4 แสดงดังตาราง 4-6

ตาราง 4-6 ตัวแปร นิยาม และกลุ่มระดับตัวแปร (State)

ลำดับ	ตัวแปรและระดับ	นิยาม
1	**Age**	กลุ่มอายุ
	15-34	อายุระหว่าง 15-34 ปี
	35-59	อายุระหว่าง 35-59 ปี
2	**Area of residence**	เขตที่พักอาศัย
	Urban	ในเขตเทศบาล
	Rural	นอกเขตเทศบาล
3	**Hypertension**	ภาวะการเป็นโรคความดันโลหิตสูง
	Normal	120 mmHg $\leq$ SBP/ 80 mmHg $\leq$ DBP
	Hypertensive	SBP < 120 mmHg/ DBP < 80 mmHg
4	**Family History HT**	ประวัติการเป็นโรคความดันโลหิตสูงของครอบครัว (สายตรง)
	Yes	มีญาติสายตรงได้แก่ พ่อ แม่ หรือพี่น้องมีประวัติเป็นโรคความดันโลหิตสูง
	No	ไม่มีญาติสายตรงได้แก่ พ่อ แม่ หรือพี่น้องมีประวัติเป็นโรคความดันโลหิตสูง
5	**Fruit and Vegetable consumption หรือ Veg**	การบริโภคผักและผลไม้
	Less	ผักผลไม้ไม่เพียงพอ < 5 servings/day
	Normal to high	ผักผลไม้เพียงพอ $\geq$ 5 servings/day
6	**Obesity**	ภาวะโรคอ้วน
	Normal	ทั้ง BMI และรอบเอวต่ำกว่าเกณฑ์
	Over BMI	BMI $\geq$ 25 (รอบเอวไม่เกินเกณฑ์)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

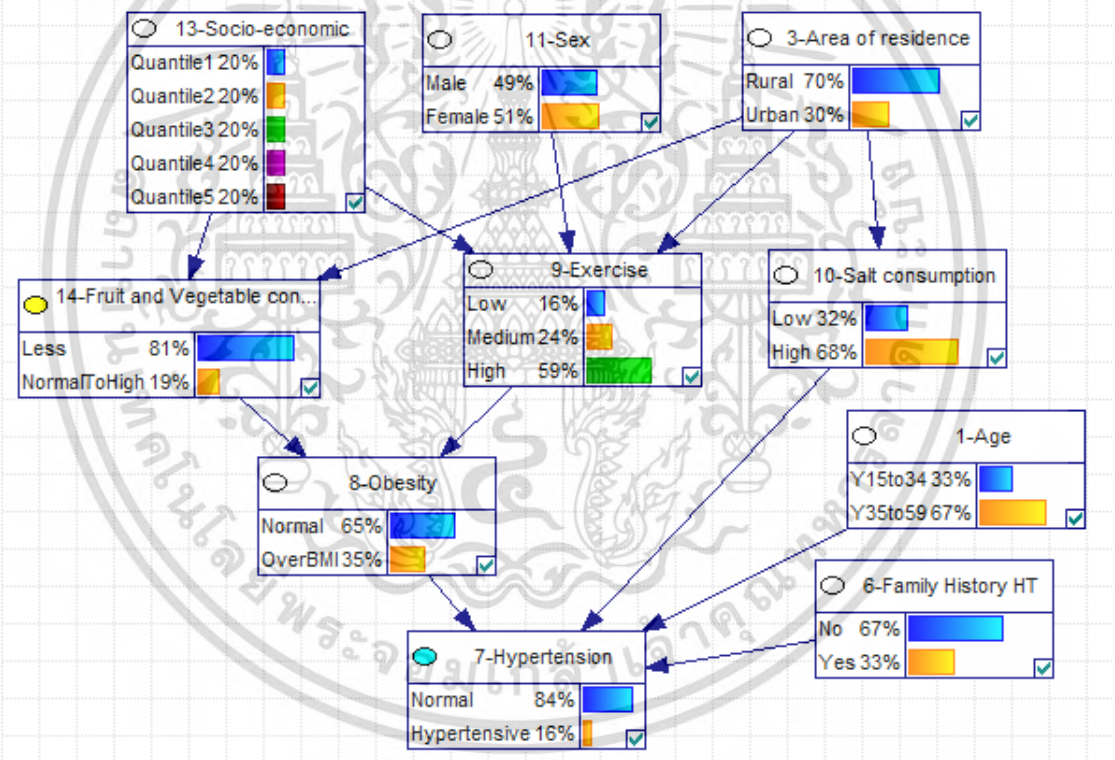
ตาราง 4-6 (ต่อ) ตัวแปร นิยาม และกลุ่มระดับตัวแปร (State)

ลำดับ	ตัวแปรและระดับ	นิยาม
7	**Socio-economic**	ความมั่นคงทางเศรษฐกิจและสังคม
	Quintile1	ตัวแปรดัชนี Wealth index score คำนวณ อยู่ใน Quintile 1
	Quintile2	ตัวแปรดัชนี Wealth index score คำนวณ อยู่ใน Quintile 2
	Quintile3	ตัวแปรดัชนี Wealth index score คำนวณ อยู่ใน Quintile 3
	Quintile4	ตัวแปรดัชนี Wealth index score คำนวณ อยู่ใน Quintile 4
	Quintile5	ตัวแปรดัชนี Wealth index score คำนวณ อยู่ใน Quintile 5
8	**Physical activity**	การเคลื่อนไหวร่างกายที่มีการใช้พลังงานในร่างกายโดยคำนึงถึงระดับความหนักเบาของกิจกรรมทางกาย
	Low	ระดับของการมีกิจกรรมทางกายต่ำกว่าเกณฑ์ระดับปานกลางและระดับมาก
	Medium	a. มีกิจกรรมทางกายไม่มากถึงระดับมาก และ a. มีกิจกรรมอย่างหนัก $\geq$ 3 วัน/สัปดาห์ และเวลา $\geq$ 20 นาทีต่อวัน b. มีกิจกรรมปานกลางหรือเดิน รวม $\geq$ 5 วัน อย่างน้อยวันละ 30 นาทีต่อวัน และ total MET-นาทีต่อสัปดาห์ $\geq$ 1,500 หรือ c. มีกิจกรรมทางกายอย่างหนักและปานกลางหรือเดิน รวม $\geq$ 5 วัน/สัปดาห์ และ total MET-นาทีต่อสัปดาห์ $\geq$ 600
	High	มีกิจกรรมทางกายอย่างหนัก $\geq$ 3 วัน/สัปดาห์ และ total MET-นาทีต่อสัปดาห์ $\geq$ 1,500 หรือ มีกิจกรรมทางกายอย่างหนักหรือปานกลางรวม $\geq$ 7 วัน/สัปดาห์ และ total MET-นาทีต่อสัปดาห์ $\geq$ 3,000
9	**Sex**	เพศ
	Male	ชาย
	Female	หญิง
10	**Salt consumption**	การบริโภคเกลือและโซเดียม
	Low	ระดับต่ำ บริโภคโซเดียม $\leq$ 2400 mg/day
	High	ระดับสูง บริโภคโซเดียม $>$ 2400 mg/day

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 โครงสร้างโมเดลเครือข่ายแบบเบย์ของการเกิดโรคความดันโลหิตสูง

การสร้างโมเดลเครือข่ายแบบเบย์แสดงความสัมพันธ์ของสาเหตุการทำให้เกิดโรคความดันโลหิตสูงจากความคิดเห็นแพทย์ที่เข้าร่วมในโครงการวิจัยในฐานนะผู้เชี่ยวชาญ (Experts) และแพทย์ผู้ดำเนินการโครงการเก็บข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย ครั้งที่ 4 ร่วมกำหนดโครงสร้างในเชิงเหตุและผลเพื่อแสดงสาเหตุของการเกิดโรคความดันโลหิตสูง โมเดลผ่านการตรวจสอบและปรับปรุงภายหลายครั้งเพื่อกำหนดเฉพาะความสัมพันธ์ที่ปรากฏด้วยลูกศรที่เชื่อมจากตัวแปรหนึ่งไปยังอีกตัวแปรต้องเป็นการเชื่อมความสัมพันธ์ที่เป็นสาเหตุโดยตรงเท่านั้น (Direct effect) จนกระทั่งได้รูปแบบโครงสร้างของโมเดลเครือข่ายแบบเบย์ที่เหมาะสมและมีข้อมูลจากการสำรวจฯ สนับสนุนการคำนวณค่าความน่าจะเป็นจนกระทั่งได้โครงสร้างโมเดล ซึ่งแสดงได้ดังต่อไปนี้



รูป 4-6 โครงสร้างโมเดลเครือข่ายแบบเบย์สำหรับการเกิดโรคความดันโลหิตสูง

จากรูปเป็นความสัมพันธ์ของตัวแปรทั้ง 10 ตัวแปรที่มีความสัมพันธ์ทั้งทางตรงและทางอ้อมไปยังการเกิดโรคความดันโลหิตสูง (Hypertension)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.3 การวิเคราะห์ปัจจัยเสี่ยงหรือตัวแปรบ่งชี้ความเสี่ยงที่สำคัญของการเกิดโรคความดันโลหิตสูงของคนไทย

ในหัวข้อนี้จะทำการวิเคราะห์เช่นเดียวการเกิดโรคเบาหวาน จากโมเดลเครือข่ายแบบเบย์ (Bayesian Network Model) เพื่อจัดลำดับปัจจัยเสี่ยงการเกิดโรคความดันโลหิตสูงในระดับประชากรไทย ในการวิเคราะห์เริ่มต้นด้วยการเลือกพิจารณาสถานะ (State) ที่สนใจของแต่ละตัวแปรสาเหตุความเสี่ยง โดยจากการพิจารณาสถานะ (State) ที่ส่งผลต่อการเกิดโรคเบาหวานในแต่ละตัวแปรกำหนดได้ดังตาราง 4-7

ตาราง 4-7 สถานะของแต่ละตัวแปรที่ใช้ในการวิเคราะห์

ตัวแปร	ชื่อตัวแปร (Variable Name)	สถานะที่สนใจ (State)
กลุ่มอายุ	Age	35-59
ประวัติการเป็นโรคความดันโลหิตสูงของครอบครัว (สายตรง)	Family History HT	Yes
เขตที่พักอาศัย	Area of residence	Rural
ความมั่นคงทางเศรษฐกิจและสังคม	Socio-economic	Quintile 4
การบริโภคผักและผลไม้	Fruit and Vegetable consumption	Normal to high
ภาวะโรคอ้วน	Obesity	Over BMI
การเคลื่อนไหวร่างกาย	Physical activity	Medium
เพศ	Sex	Female
การบริโภคเกลือและโซเดียม	Salt consumption	High

งานวิจัยนี้นำเสนอการจัดลำดับปัจจัยเสี่ยง (Risk Prioritisation) ต่อการเกิดโรคความดันโลหิตสูง การจัดกลุ่มปัจจัยเสี่ยง (Risk Classification) ที่อาจส่งผลต่อการเกิดความดันโลหิตสูง เช่นเดียวกับการวิเคราะห์การเกิดโรคเบาหวาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.3.1 การจัดลำดับความเสี่ยง (Risk Prioritisation) ต่อการเกิดโรคความดันโลหิตสูงของคนไทย

การวิเคราะห์จากโมเดลเครือข่ายแบบเบย์ (Bayesian Network Model) ของการเกิดโรคความดันโลหิตสูง เพื่อหาคำตอบของคำถามที่ว่า “อะไรเป็นปัจจัยหลักที่ส่งผลต่อการเป็นโรคความดันโลหิตสูง?”

คำนวณ Normalised Likelihood ซึ่งเป็นอัตราส่วนระหว่างความน่าจะเป็นที่เปลี่ยนไปของแต่ละปัจจัยเมื่อทราบว่าเป็นผู้ป่วยโรคเบาหวาน (Adjusted Probability) และความน่าจะเป็นที่เกิดขึ้นในปัจจุบัน (Current Probability) โดยทั่วไปในสมการ (2-7) ดังนั้นเราจะคำนวณค่า Normalised Likelihood และเปอร์เซ็นต์ที่เพิ่มขึ้น และกำหนดค่าดังกล่าวในการเรียงลำดับปัจจัยเสี่ยงที่เกี่ยวข้องต่อการเกิดโรคความดันโลหิตสูง ของทั้ง 10 ตัวแปรแสดงได้ดังตาราง 4.8

ตาราง 4-8 ผลการเรียงลำดับปัจจัยเสี่ยงต่อการเกิดโรคความดันโลหิตสูงของประชาชนไทย

ชื่อตัวแปร (Variable Name)	สถานะที่สนใจ (State)	ความน่าจะเป็น		Normalised likelihood	% of Increase	ลำดับ (Rank)
		ปัจจุบัน (Current)	เปลี่ยนไป (Adjusted)			
Age	35-59	0.6664	0.8864	1.3302	33.0167	3
Family History HT	Yes	0.3285	0.4377	1.3324	33.2438	2
Area of residence	Rural	0.6955	0.6995	1.0057	0.5684	7
Socio-economic	Quintile 4	0.2	0.2002	1.0010	0.0994	8
Fruit and Vegetable consumption	Normal to high	0.1857	0.1897	1.0214	2.1410	4
Obesity	Over BMI	0.3547	0.5251	1.4804	48.0449	1
Physical activity	Medium	0.2438	0.2475	1.0154	1.5411	5
Sex	Female	0.5057	0.5061	1.0007	0.0718	9
Salt consumption	High	0.6815	0.6883	1.0101	1.0052	6
Hypertension	Positive	0.1584				

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างเช่น โอกาสที่จะพบผู้มีภาวะโรคอ้วน (Obesity) เพิ่มขึ้นจาก 0.3547 เมื่อเป็นกลุ่มของประชาชนไทยทั่วไป เป็น 0.5251 เมื่อทราบว่าเขาเป็นโรคความดันโลหิตสูง ทำให้มีอัตราส่วนที่เพิ่มขึ้น เป็น 1.4804 เท่า คิดเป็นโอกาสที่เพิ่มขึ้น 48.0449%

จากตารางข้างต้นพบว่าปัจจัยด้านภาวะโรคอ้วน (Obesity) เป็นปัจจัยที่มีอัตราส่วนการเปลี่ยนแปลงความน่าจะเป็นเพิ่มขึ้นเมื่อทราบว่าเป็นโรคเบาหวานเทียบกับความน่าจะเป็นที่เกิดขึ้นทั่วไปคิดเป็นร้อยละ 48.0449 รองลงมาคือ ประวัติการเป็นโรคความดันโลหิตสูงของครอบครัวสายตรง (Family History HT) มีความน่าจะเป็นเพิ่มขึ้นคิดเป็นร้อยละ 33.2438 ปัจจัยเสี่ยงอันดับ 3 ที่มีค่าใกล้เคียงกับอันดับ 2 ได้แก่ปัจจัยด้านอายุ (Age) ในกลุ่มอายุ 35-59 ปี มีความน่าจะเป็นเพิ่มขึ้นคิดเป็นร้อยละ 33.0167 สำหรับปัจจัยตั้งแต่ลำดับที่ 4 เป็นต้นไป มีความน่าจะเป็นเพิ่มขึ้นค่อนข้างต่ำ โดยพบว่า ปัจจัยเสี่ยงอันดับ 4 ได้แก่ปัจจัยด้านการบริโภคผักและผลไม้ (Fruit and Vegetable Consumption) ในระดับปานกลางถึงมากเพิ่มขึ้น (เพิ่มขึ้น 2.1410%) ปัจจัยเสี่ยงอันดับ 5 เคลื่อนไหวทางกาย (Physical Activity) ในระดับปานกลาง (เพิ่มขึ้น1.5411%) ปัจจัยเสี่ยงอันดับ 6 ได้แก่การบริโภคเกลือและโซเดียม ในระดับสูง (เพิ่มขึ้น 1.0052%) ปัจจัยเสี่ยงอันดับ 7 คือปัจจัยด้านที่อยู่อาศัย (Area of Residence) นอกเขตเทศบาล (เพิ่มขึ้น 0.5684%) ปัจจัยเสี่ยงอันดับ 8 สถานะทางสังคม (Socio-Economic) ในระดับ Quintile 4 (เพิ่มขึ้น 0.0994%) ปัจจัยเสี่ยงอันดับสุดท้ายคือเพศ ในกลุ่มเพศหญิง (0.0718%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.3.2 การจัดกลุ่มปัจจัยเสี่ยง (Risk classification) ที่อาจส่งผลต่อการเกิดโรคความดันโลหิตสูงของคนไทย

การวิเคราะห์จากโมเดลเครือข่ายแบบเบย์ (Bayesian Network Model) ของการเกิดโรคความดันโลหิตสูงเพื่อหาคำตอบของคำถามที่ว่า “ปัจจัยความเสี่ยงการเกิดโรคความดันโลหิตสูงของคนไทยควรจัดการอย่างไร”

การกำหนดแนวทางการจัดการสาเหตุการเกิดโรคเบาหวานใช้แนวทางการพิจารณาจากการกำหนดตารางประเมินความเสี่ยง (Risk Matrix) โดยแบ่งกลุ่มสาเหตุความเสี่ยงของการเกิดโรคจาก 2 องค์ประกอบหลักได้แก่ ความน่าจะเป็นของการเกิดความเสี่ยงของแต่ละสาเหตุการเกิดโรคซึ่งในที่นี้เรียกปัจจัยเสี่ยงซึ่งนำเสนอในรูปความน่าจะเป็นที่เกิดขึ้นในปัจจุบัน (Current Probability) โดยทั่วไปและผลกระทบที่เกิดจากความน่าจะเป็นที่เปลี่ยนไปเมื่อทราบว่าเป็นผู้ป่วยโรคเบาหวานที่นำเสนอในรูปของ % ของความน่าจะเป็นที่เพิ่มขึ้นของแต่ละปัจจัยเสี่ยงเมื่อทราบว่าเป็นผู้ป่วยความดันโลหิตสูง (% Increase) ทั้งสองปัจจัยเป็นตัวกำหนดแกน X และ Y

พื้นที่ในกราฟสามารถแบ่งได้เป็น 4 ควอดรันต์ (Quadrant) หรือ จตุภาค ตามการแบ่งค่าสูงและต่ำของทั้ง 2 ปัจจัยดังกล่าว จากนั้นเลือกแต่ละปัจจัยที่มี State ที่ให้ค่า Normalised Likelihood มากกว่า 1 เป็นตัวแทนแต่ละปัจจัยเพื่อนำมาแบ่งกลุ่ม (ดังที่ได้แสดงดังตาราง 4-7) แสดงผลการแบ่งกลุ่มปัจจัยเสี่ยงได้ดังรูป 4-6 จากรูปจะเห็นได้ว่าหากกำหนดจุดตัดเพื่อแบ่งค่าสูงและต่ำที่ 50% ซึ่งประกอบด้วย กลุ่ม High Probability & High Impact กลุ่ม Low Probability & High Impact กลุ่ม High Probability & Low Impact กลุ่ม Low Probability & Low Impact โดยพบว่าไม่มีปัจจัยใดอยู่ในกลุ่ม High Probability & High Impact ส่วนในกลุ่มที่เหลือมีผลได้ดังนี้

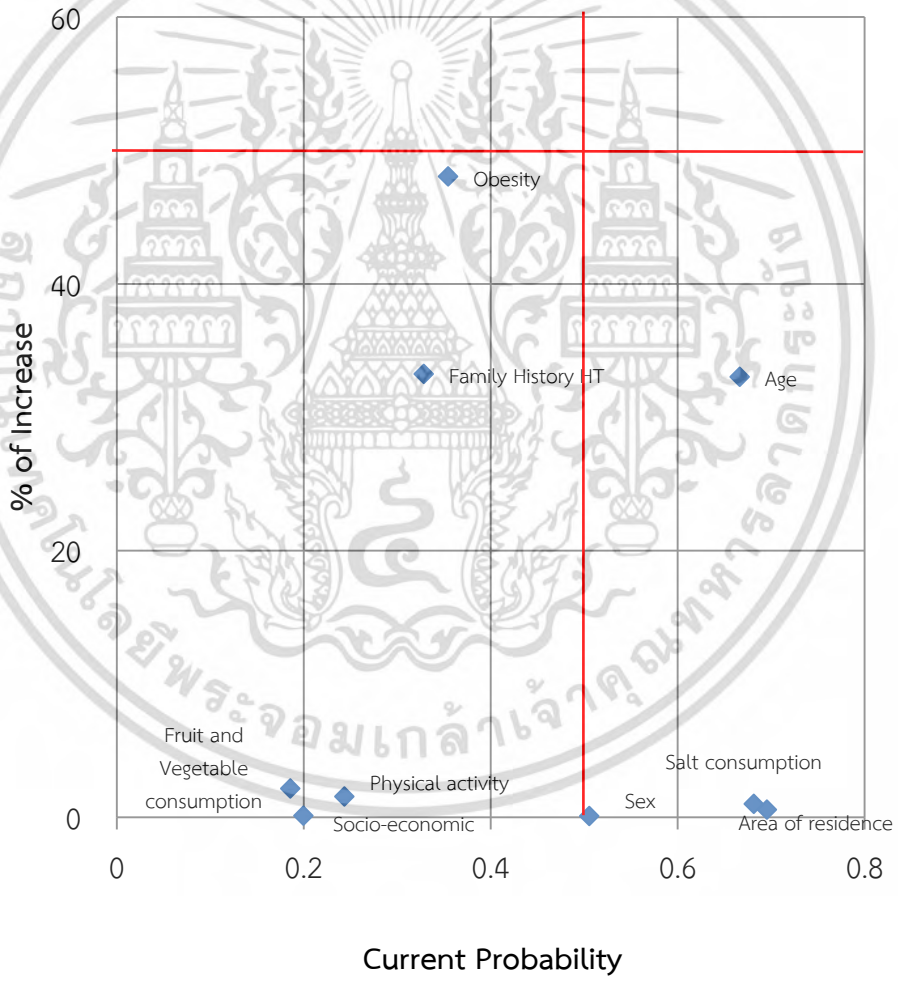
ในที่นี้หากกำหนดจุดแบ่งระหว่างระดับสูงและต่ำของปัจจัยส่งผลกระทบต่อการเกิดโรคความดันโลหิตสูง (Impact) ที่แสดงโดยค่าความน่าจะเป็นที่เพิ่มขึ้นของแต่ละปัจจัยเสี่ยงเมื่อทราบว่าเป็นผู้ป่วยความดันโลหิตสูง (% Increase) ที่ระดับ 50% พบว่าทุกตัวแปรมีผลกระทบต่อการเกิดโรคความดันโลหิตสูงอยู่ในระดับต่ำ เมื่อพิจารณาแบ่งความน่าจะเป็นของการเกิดแต่ละปัจจัยในปัจจุบัน (Current Probability) ที่ 0.5 พบว่ามีตัวแปรที่มีโอกาสเกิดทั้งในระดับสูง (High Prob) และระดับต่ำ (Low Prob) นั้นคือสามารถแบ่งกลุ่มตัวแปรความเสี่ยงได้เป็น 2 กลุ่มคือ High Probability & Low Impact กลุ่ม Low Probability & Low Impact ดังนี้

กลุ่ม Low Probability & Low Impact ประกอบด้วย 5 ปัจจัย ดังนี้ปัจจัยด้านภาวะโรคอ้วน (Obesity) ปัจจัยด้านประวัติการเป็นโรคความดันโลหิตสูงของครอบครัวสายตรง (Family History Ht) ปัจจัยด้านการบริโภคผักและผลไม้ (Fruit And Vegetable Consumption) ในระดับปานกลางถึงมาก ปัจจัยด้านเคลื่อนไหวทางกาย (Physical Activity) ปัจจัยด้านสถานะทางสังคม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Socio-Economic) ในระดับ Quintile 4 การจัดการความเสี่ยงจะเลือกวิธีการ Retain/ Absorption โดยอาจไม่ต้องให้ความสำคัญกับปัจจัยเหล่านี้

กลุ่ม High Probability & Low Impact ประกอบด้วย 4 ปัจจัย ได้แก่ปัจจัยด้านอายุ (Age) มี ในกลุ่มอายุ 35-59 ปี ปัจจัยด้านการบริโภคเกลือและโซเดียม (Salt Consumption) ในระดับสูง ปัจจัยด้านที่อยู่อาศัย (Area Of Residence) นอกเขตเทศบาล และสุดท้ายคือปัจจัยด้านเพศ (Sex) ในกลุ่มเพศหญิง การจัดการความเสี่ยงจะเลือกวิธีการ Reduction/Mitigation เช่น การจัดการรณรงค์ให้ลดอาหารที่มีรดชาติเค็มโดยเฉพาะในกลุ่มเพศหญิง อาศัยอยู่ในเขตเทศบาล ที่มีอายุ 35-59 ปี หรือให้มีการตรวจคัดกรองการเกิดโรคความดันโลหิตสูง



รูป 4-7 ผลการจัดกลุ่มความเสี่ยงการเกิดโรคความดันโลหิตสูง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.4 พยากรณ์การเกิดโรคความดันโลหิตสูงของคนไทย

การสร้างตัวแบบเพื่อใช้พยากรณ์การเกิดโรคเบาหวานด้วยโมเดลเครือข่ายแบบเบย์ จากจำนวนหน่วยหน่วยตัวอย่างที่มีอายุ 15-59 ปี จำนวนทั้งสิ้น 11,240 คน โดยดำเนินการโดยใช้ตัวแบบที่สร้างขึ้นจากข้อมูลที่ใช้ในการสร้างตัวแบบ (Training Dataset) และสำหรับการทดสอบความสามารถของโมเดล (Testing Dataset) โดยใช้อัตราส่วน 70:30 การแบ่งข้อมูลเป็น 2 กลุ่มเกิดขึ้นอย่างสุ่มซึ่งเป็นกลุ่มเดียวกับที่นำมาวิเคราะห์การเกิดโรคเบาหวาน ในหัวข้อ 4.1.4 โดยจะนำเสนอผลการวิเคราะห์ในหัวข้อนี้จากชุดข้อมูลจากการทดสอบ (Testing Dataset) เป็นหลัก เพื่อให้เชื่อมั่นได้ว่าโมเดลมีความสามารถในการจำแนกกลุ่มการพยากรณ์การเกิดโรคความดันโลหิตสูงหรือไม่ (Positive Or Negative) ได้อย่างถูกต้องมากน้อยเพียงใด แสดงผลการตรวจสอบประสิทธิภาพการพยากรณ์จาก คอนฟูชั่น แมททริก คะแนน F1 และเส้นโค้ง ROC

ผลจากตาราง 4-9 เป็นผลจากการใช้โมเดลในการจำแนกกลุ่มจากข้อมูลการทดสอบ (Testing Dataset) ทำให้ทราบว่าโมเดลเครือข่ายแบบเบย์มีความผิดพลาดในการจำแนกกลุ่มการเกิดโรคความดันโลหิตสูง ระหว่าง 18.96% – 53.41% โดยมีค่าความผิดพลาดเพิ่มขึ้นเมื่อกำหนดระดับ Threshold ลดลง และพบว่าที่ระดับ Threshold = 0.5, 0.4, 0.3 จะให้ค่าความผิดพลาด (Error Rate) 18.96% เพิ่มขึ้นเป็น 20.52% ที่ระดับเกณฑ์ (Threshold) = 0.3 และเพิ่มขึ้นเป็น 34.53% ที่ระดับ Threshold = 0.2 และเป็น 53.41% ระดับ Threshold = 0.1 หรืออาจกล่าวได้ว่าค่าความถูกต้อง (Accuracy) ลดลงจาก 81.04% เป็น 46.59% เมื่อกำหนดลดระดับเกณฑ์ Threshold ลดลง

นอกจากนี้ยังพบว่าโมเดลเครือข่ายแบบเบย์ให้ค่าความจำเพาะ (Specificity) และค่าทำนายเมื่อผลเป็นลบ (NPV) ในระดับสูง และยังพบว่าโมเดลจะให้ค่าความจำเพาะ (Specificity) และค่าความแม่นยำ (Precision) มีแนวโน้มลดลงเมื่อมีระดับเกณฑ์ (Threshold) ลดลง ในขณะที่โมเดลจะให้ค่าทำนายเมื่อผลเป็นลบ (NPV) และค่าความระลึก (Recall) เพิ่มขึ้นเมื่อมีระดับเกณฑ์ (Threshold) ลดลง

จะเห็นได้จากโมเดลเมื่อลดระดับเกณฑ์ (Threshold) ลงจะทำให้มีค่า True Positive ค่อนต่ำลงแต่ค่า True Negative สูงขึ้น ส่งผลให้ค่าที่ใช้เป็นเกณฑ์การพิจารณาประสิทธิภาพของโมเดลไม่เป็นไปในทิศทางเดียวกัน เช่นหากมีค่าความระลึก (Recall) ลดลง แต่ค่าความแม่นยำ (Precision) เพิ่มขึ้นเมื่อลดระดับเกณฑ์ (Threshold) ลดลงเป็นต้น เนื่องจากทั้ง 2 ค่า ดังกล่าวมีทิศทางความสัมพันธ์ที่ขัดแย้งกัน จึงมักใช้ค่า F1 เพื่อหาค่าเฉลี่ยของทั้ง 2 ค่า และพบว่าค่า F1 มีค่าสูงสุดเมื่อกำหนดระดับเกณฑ์ (Threshold) เป็น 0.2 ทำให้มีค่า F1 เป็น 0.4412

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 4-9 คอนฟูชั่น แมททริก และค่าต่างๆ ที่ใช้ในการวัดความสามารถในการจำแนกโรคความดันโลหิตสูง ในระดับเกณฑ์ความน่าจะเป็นที่แตกต่างกัน

ค่าจริง (Actual)	Threshold level									
	0.5		0.4		0.3		0.2		0.1	
	ค่าพยากรณ์ (Predictive)									
	P	N	P	N	P	N	P	N	P	N
BNH model										
P	0	632	0	632	175	457	452	180	587	45
N	0	2,701	0	2,701	227	2,474	965	1,736	1,735	966
Error rate ความผิดพลาด	18.96%		18.96%		20.52%		34.53%		53.41%	
ความถูกต้อง (Accuracy)	0.8104		0.8104		0.7948		0.6565		0.4659	
ค่าความจำเพาะ (Specificity)	1.0000		1.0000		0.9160		0.6427		0.3576	
ค่าทำนายเมื่อผลเป็นลบ (NPV)	0.8104		0.8104		0.8441		0.9061		0.9555	
ความระลึก (Recall)	0.0000		0.0000		0.2769		0.7152		0.9288	
ค่าความแม่นยำ (Precision)	-		-		0.4353		0.3190		0.2528	
F1 score	-		-		0.3385		0.4412		0.3974	

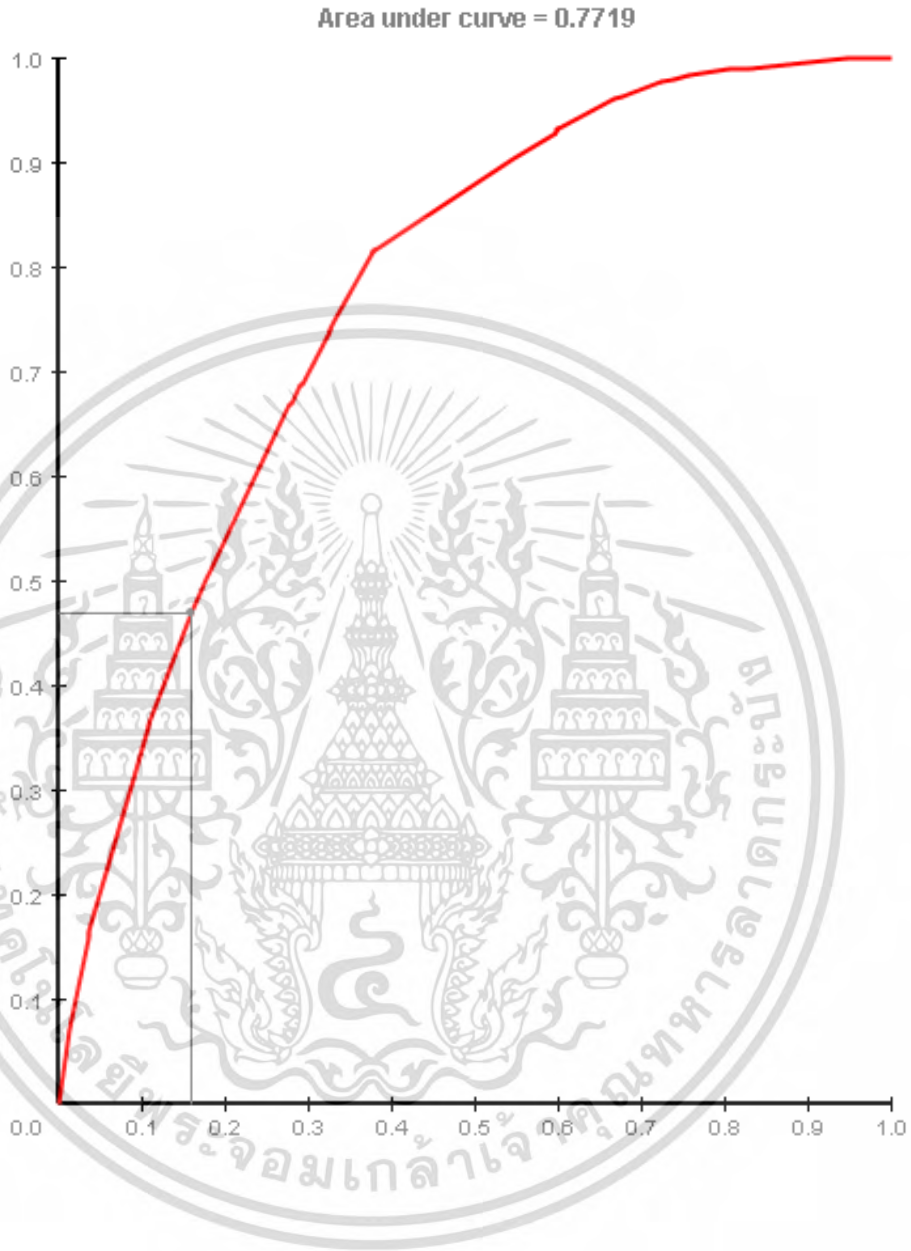
P: Positive; N: Negative

ตาราง 4-10 ค่า AUC จาก Training และ Testing dataset และ Full Dataset สำหรับโรคความดันโลหิตสูง

	Training Dataset	Testing Dataset	Full Dataset
AUC	0.7719	0.7371	0.7418

จากตาราง 4-10 แสดงค่าพื้นที่ใต้โค้ง ROC ความสามารถในการทำนายผลอย่างถูกต้องจากโมเดลเทียบกับค่าจริงที่ได้จากชุดของข้อมูลที่นำมาศึกษา โดยพบว่า มีเป็น 77.19 และ 73.71 ซึ่งคำนวณจาก Training และ Testing dataset ตามลำดับ หากพิจารณาค่า AUC ของข้อมูลปกติที่ไม่ได้แบ่งเป็น Training และ Testing Dataset พบว่าสามารถทายถูก 74.188 ซึ่งเป็นค่าที่อยู่ระหว่างค่าที่ได้จากชุดข้อมูล Training และ Testing Dataset

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

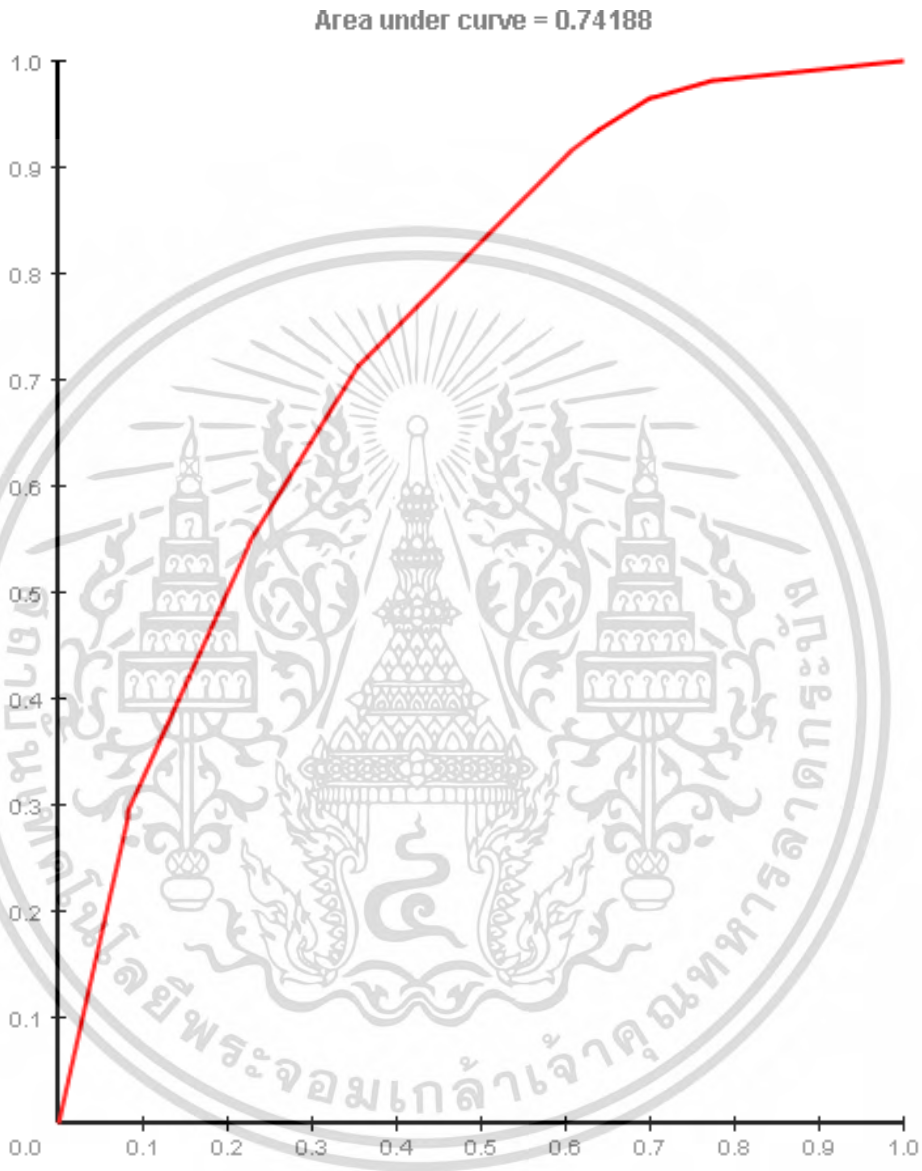


รูป 4-8 ROC Curve จาก Training Dataset สำหรับโรคความดันโลหิตสูง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูป 4-9 ROC Curve จาก Testing Dataset สำหรับโรคความดันโลหิตสูง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 4-10 ROC Curve จาก Full Dataset สำหรับโรคความดันโลหิตสูง

เมื่อพิจารณาเส้นโค้ง ROC ดังรูป 4-3-4-5 กราฟเป็นเส้นโค้ง ROC มีลักษณะโค้งเข้ามุมบนซ้ายมือ ที่ให้พื้นที่ใต้โค้งใกล้เคียงกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5 สรุปผลการวิจัยและอภิปราย

การศึกษาครั้งนี้เป็นการสร้างโมเดลเครือข่ายแบบเบย์ เพื่อแสดงปัจจัยเชิงเหตุและผลต่อการเกิดโรคเบาหวานและโรคความดันโลหิตสูง ซึ่งเป็นโรคที่พบมากในกลุ่มโรคที่ไม่ติดต่อเรื้อรัง ผู้วิจัยได้ทำการสร้างโมเดลเครือข่ายแบบเบย์ด้วยวิธีการสัมภาษณ์กับผู้เชี่ยวชาญหลายครั้งก่อนกำหนดโครงสร้างของโมเดลของโรคเบาหวานที่ประกอบด้วยตัวแปรทั้งสิ้น 8 ตัวแปร โครงสร้างของโมเดลของโรคความดันโลหิตสูงที่ประกอบด้วยตัวแปรทั้งสิ้น 10 ตัวแปร และทำการกำหนดค่าความน่าจะเป็นและความน่าจะเป็นแบบมีเงื่อนไขของแต่ละตัวแปรโดยอาศัยข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย ครั้งที่ 4 ก่อนทำการวิเคราะห์ตามแนวทางที่งานวิจัยนี้เสนอเพื่อวัตถุประสงค์ของงานวิจัยนี้ 2 ประการ คือ

1. วิเคราะห์ปัจจัยเสี่ยงหรือตัวบ่งชี้ความเสี่ยงที่สำคัญต่อการเกิดโรคไม่ติดต่อที่สำคัญของคนไทยจากข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกายจากโมเดลเครือข่ายแบบเบย์ (Bayesian Network Model) โดยวิธีจัดลำดับปัจจัยเสี่ยง และจัดกลุ่มความเสี่ยง
2. พยากรณ์การเกิดโรคไม่ติดต่อที่สำคัญของคนไทยจากข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกายจากโมเดลเครือข่ายแบบเบย์ (Bayesian Network Model)

ผลจากการวิเคราะห์ดังกล่าวสามารถสรุปได้ดังหัวข้อที่ 5.1 นอกจากนี้ยังมีการสรุปผลการวิเคราะห์เหตุผลที่การนำเสนอตัวแบบเครือข่ายแบบเบย์ในการวิเคราะห์การเกิดโรคไม่ติดต่อในหัวข้อ 5.2 และข้อเสนอแนะในหัวข้อ 5.3

5.1 ผลจากการวิเคราะห์ด้วยโมเดลเครือข่ายแบบเบย์

5.1.1 การจัดลำดับความเสี่ยง

การจัดลำดับความเสี่ยงของตัวแปรที่ส่งผลต่อการเกิดโรคเบาหวานสำหรับงานวิจัยนี้ เพื่อตอบคำถามว่า “อะไรเป็นปัจจัยหลักที่ส่งผลต่อการเป็นโรคเบาหวาน/ความดันโลหิตสูง?” โดยการเปรียบเทียบคำนวณค่า Normalised Likelihood และ เปอร์เซ็นต์ที่เพิ่มขึ้นของแต่ละตัวแปร สำหรับโรคเบาหวาน พบว่าปัจจัยด้านประวัติการเป็นโรคเบาหวานของครอบครัวสายตรง (Family History DM) เป็นความเสี่ยงในลำดับที่ 1 รองลงมาคือการเป็นโรคอ้วน (Obesity) ปัจจัยเสี่ยงอันดับ 3 ได้แก่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปัจจัยด้านอายุ (Age) ในกลุ่มอายุ 35-59 ปี ปัจจัยเสี่ยงอันดับ 4 ได้แก่ปัจจัยด้านการเคลื่อนไหวทางกาย (Physical Activity) ในระดับปานกลาง ปัจจัยเสี่ยงอันดับ 5 ได้แก่สถานะทางสังคม (Socio-Economic) ในระดับ Quintile 4 ปัจจัยเสี่ยงอันดับ 6 ได้แก่การบริโภคผักและผลไม้ (Fruit and Vegetable Consumption) ในระดับน้อย และปัจจัยเสี่ยงอันดับสุดท้ายคือด้านที่อยู่อาศัย (Area of Residence) ในเขตเทศบาล

สำหรับโรคความดันโลหิตสูง พบว่าปัจจัยด้านภาวะโรคอ้วน (Obesity) เป็นปัจจัยที่มีเป็นความเสี่ยงในลำดับที่ 1 รองลงมาคือ ประวัติการเป็นโรคความดันโลหิตสูงของครอบครัวสายตรง (Family History HT) ปัจจัยเสี่ยงอันดับ 2 และปัจจัยด้านอายุ (Age) ในกลุ่มอายุ 35-59 ปี เป็นปัจจัยเสี่ยงอันดับ 3 สำหรับลำดับที่ 4 เป็นต้นไป มีระดับความเสี่ยงต่ำโดยพบว่าปัจจัยเสี่ยงอันดับ 4 ได้แก่ปัจจัยด้านการบริโภคผักและผลไม้ (Fruit and Vegetable Consumption) ในระดับปานกลางถึงมาก ปัจจัยเสี่ยงอันดับ 5 เคลื่อนไหวทางกาย (Physical Activity) ปัจจัยเสี่ยงอันดับ 6 ได้แก่การบริโภคเกลือและโซเดียมในระดับสูง ปัจจัยเสี่ยงอันดับ 7 คือปัจจัยด้านที่อยู่อาศัย (Area of Residence) นอกเขตเทศบาล ปัจจัยเสี่ยงอันดับ 8 สถานะทางสังคม (Socio-Economic) ในระดับ Quintile 4 ปัจจัยเสี่ยงอันดับสุดท้ายคือเพศ ในกลุ่มเพศหญิง (0.0718%)

ทั้ง 3 ปัจจัยหลักของการเกิดโรคเบาหวานและความดันโลหิตสูง คือกลุ่มปัจจัยกลุ่มเดียวกัน แต่อาจมีลำดับเปลี่ยนแปลงสลับกันเล็กน้อย เช่นประวัติการเป็นโรคเบาหวานของครอบครัวสายตรง (Family History DM) เป็นความเสี่ยงในลำดับที่ 1 ในโมเดลโรคเบาหวาน และปัจจัยด้านภาวะโรคอ้วน (Obesity) เป็นปัจจัยที่มีเป็นความเสี่ยงในลำดับที่ 1 ในโมเดลโรคความดันโลหิตสูง

5.1.2 การจัดกลุ่มความเสี่ยง

การจัดกลุ่มความเสี่ยงเพื่อให้ผู้วางแผนนโยบายสามารถกำหนดแนวทางนโยบายในการจัดการกับความเสี่ยงของการเกิดโรคเบาหวานได้อย่างเหมาะสม นั้นคือการวิเคราะห์นี้เพื่อหาคำตอบของคำถามที่ว่า “ปัจจัยความเสี่ยงการเกิดโรคเบาหวาน/ความดันโลหิตสูงของคนไทยควรจัดการอย่างไร”

จากการวิเคราะห์ด้วยตารางประเมินความเสี่ยง (Risk Matrix) โดยแบ่งกลุ่มสาเหตุความเสี่ยงของการเกิดโรคจาก 2 องค์ประกอบหลักได้แก่ ความน่าจะเป็นของการเกิดความเสี่ยงของแต่ละสาเหตุการเกิดโรคด้วยความน่าจะเป็นที่เกิดขึ้นในปัจจุบัน (Current Probability) โดยทั่วไป และผลกระทบที่ในรูปของ % ของความน่าจะเป็นที่เพิ่มขึ้นของแต่ละปัจจัยเสี่ยงเมื่อทราบว่าเป็นผู้ป่วยเบาหวาน/ความดันโลหิตสูง (% Increase) ทั้งสองปัจจัยเป็นตัวกำหนดแกน X และ Y โดยการกำหนดจุดตัดเพื่อแบ่งค่าสูงและต่ำที่ 50% ซึ่งประกอบด้วย กลุ่ม High Probability & High Impact

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กลุ่ม Low Probability & High Impact กลุ่ม High Probability & Low Impact กลุ่ม Low Probability & Low Impact โดยพบว่าไม่มีปัจจัยใดอยู่ในกลุ่ม High Probability & High Impact ส่วนในกลุ่มที่เหลือมีผลได้ดังนี้

กลุ่มที่ 1 Low Probability & High Impact ในกลุ่มโรคเบาหวานได้แก่ ปัจจัยที่ด้านประวัติการเป็นโรคเบาหวานของครอบครัวสายตรง (Family History DM) และปัจจัยด้านการเป็นโรคอ้วน (Obesity) สำหรับโมเดลการเกิดโรคความดันโลหิตสูง ไม่มีโมเดลใดอยู่ในกลุ่มนี้ การจัดการลดความเสี่ยงให้กลับบุคคลกลุ่มนี้ทำได้โดยการ Transfer ความเสี่ยง เช่น แนะนำให้มีการทำประกันสุขภาพที่สามารถครอบคลุมการรักษาโรคเบาหวาน

กลุ่มที่ 2 High Probability & Low Impact ในกลุ่มโรคเบาหวานประกอบด้วย 2 ปัจจัย ได้แก่ กลุ่มประชากรที่มีอายุ (Age) ในช่วง 35-59 ปี (ไม่รวมกลุ่มผู้สูงอายุ) พบว่าเป็นในกลุ่มประชากรส่วนใหญ่ (เนื่องจากสังคมไทยเข้าสู่ประชาคมสูงอายุ) และปัจจัยด้านการบริโภคผักและผลไม้ในปริมาณน้อย สำหรับกลุ่มโรคความดันโลหิตสูง ประกอบด้วย 4 ปัจจัย ได้แก่ปัจจัยด้านอายุ (Age) มีในกลุ่มอายุ 35-59 ปี ปัจจัยด้านการบริโภคเกลือและโซเดียม (Salt Consumption) ในระดับสูง ปัจจัยด้านที่อยู่อาศัย (Area Of Residence) นอกเขตเทศบาล และสุดท้ายคือปัจจัยด้านเพศ (Sex) ในกลุ่มเพศหญิง ควรการจัดการความเสี่ยงจะเลือกวิธีการ Reduction/Mitigation เช่น การจัดการรณรงค์ให้ลดทางอาหารเค็มโดยเฉพาะในกลุ่มเพศหญิง อาศัยอยู่ในเขตเทศบาล ที่มีอายุ 35-59 ปี หรือให้มีการตรวจคัดกรองการเกิดโรคความดันโลหิตสูง การจัดการความเสี่ยงจะเลือกวิธีการ Reduction/Mitigation เช่น การสนับสนุนให้คนไทยบริโภคผักและผลไม้มากขึ้น

กลุ่มที่ 3 Low Probability & Low Impact ในกลุ่มโรคเบาหวาน ประกอบด้วย 3 ตัวแปร ได้แก่ กลุ่มประชากรที่อาศัย (Area of Residence) อยู่ในเมือง หรือมีสถานะทางเศรษฐกิจและสังคม (Socio-Economic) ในระดับ 4 หรือมีกิจกรรมทางกาย (Physical Activity) ในระดับต่ำ สำหรับโมเดลโรคความดันโลหิตสูง ประกอบด้วย 5 ตัวแปร ดังนี้ ปัจจัยด้านภาวะโรคอ้วน (Obesity) ปัจจัยด้านประวัติการเป็นโรคความดันโลหิตสูงของครอบครัวสายตรง (Family History HT) ปัจจัยด้านการบริโภคผักและผลไม้ (Fruit And Vegetable Consumption) ในระดับปานกลางถึงมาก ปัจจัยด้านเคลื่อนไหวทางกาย (Physical Activity) ปัจจัยด้านสถานะทางสังคม (Socio-Economic) ในระดับ Quintile 4 การจัดการความเสี่ยงจะเลือกวิธีการ Retain/Absorption โดยผู้มีอำนาจตัดสินใจอาจไม่ต้องให้ความสำคัญกับปัจจัยเหล่านี้

กลุ่มที่ 4 High Probability & Low Impact พบว่า ไม่มีตัวแปรที่ถูกจัดกลุ่มในกลุ่มนี้ทั้งโมเดลโรคเบาหวานและความดันโลหิตสูง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.1.3 การพยากรณ์ความเสี่ยง

จากการพยากรณ์การเกิดโรคเบาหวานโดยพิจารณาจากค่า AUC พบว่า โมเดลมีความสามารถในการทำนายผลอย่างถูกต้องเป็น 76.70 และ 77.60 ซึ่งคำนวณจาก Training และ Testing Dataset ตามลำดับ จากการพยากรณ์การเกิดโรคความดันโลหิตสูงพบว่า มีค่าเป็น 77.19 และ 73.71 ซึ่งคำนวณจาก Training และ Testing dataset ตามลำดับ

ในการเปรียบเทียบความสามารถของโมเดลโดยวิธีเครือข่ายแบบเบย์ อาจพิจารณาจากค่าต่างๆ ได้หลายค่า เช่น ความถูกต้องของโมเดล (Accuracy) ค่าความไว (Sensitivity) /ความระลึก (Recall) ค่าความจำเพาะ (Precision) ซึ่งเป็นค่าที่สามารถคำนวณได้จากคอนฟิวชั่น แมททริก แต่พบว่าค่าต่างๆ มักไม่เป็นไปในทิศทางเดียวกัน หรือกล่าวได้ว่าบางโมเดลจะสลับกันให้ค่าที่ดีที่สุดไม่เหมือนกัน โดยพบว่าค่าความถูกต้องของโมเดล (Accuracy) จะใช้เมื่อพิจารณาแล้วภายใต้เงื่อนไขว่าความผิดพลาดที่เกิดขึ้นทั้ง Fault Negative และ Fault Positive มีต้นทุนที่ใกล้เคียงกัน และจะใช้ค่าคะแนน F1 ซึ่งเป็นค่าที่นำค่า Precision and Recall มาคำนวณซึ่งอาจกล่าวได้ว่าเป็นค่าที่มีประโยชน์กว่าการใช้ค่า ความถูกต้องของโมเดล (Accuracy) โดยเฉพาะอย่างยิ่งเมื่อความผิดพลาดที่เกิดขึ้นทั้ง Fault Negative และ Fault Positive มีต้นทุนที่แตกต่างกัน ดังนั้นจึงมักพบว่าจะใช้ค่าคะแนน F1 เป็นเกณฑ์ในการตัดสิน

อย่างไรก็ตามค่าคะแนน F1 สามารถหาได้กรณีที่สามารถหาค่าความแม่นยำได้ (Precision) ได้เท่านั้นซึ่งหมายถึงโมเดลจะต้องสามารถจำแนกกลุ่ม ผลบวกจริง (True Positives: TP) ซึ่งจากการใช้โมเดลเครือข่ายแบบเบย์จะให้ผลการจำแนกกลุ่มผลบวกจริง (True Positives: TP) ที่ไม่เป็น 0 เมื่อกำหนดในระดับเกณฑ์ความน่าจะเป็น (Threshold level) ที่ต่ำกว่า 0.5

ในกรณีที่คิดว่าต้นทุนของ Fault Negative และ Fault Positive ใกล้เคียงกัน การใช้โมเดลเครือข่ายแบบเบย์สำหรับโรคเบาหวานจะมีความถูกต้องสูงสุดเมื่อ ในกรณีที่กำหนดเกณฑ์ (Threshold) เป็น 0.3-0.5 โดยให้ค่าความถูกต้องของโมเดล (Accuracy) เป็น 94.32% ซึ่งเป็นค่าค่อนข้างสูง เมื่อพิจารณาผลการพยากรณ์ของโมเดลเครือข่ายแบบเบย์ สำหรับโรคความดันโลหิตสูง พบว่า ในกรณีที่กำหนดเกณฑ์ (Threshold) เป็น 0.4-0.5 โดยให้ค่าความถูกต้องของโมเดล (Accuracy) เป็น 81.04%

แต่หากเชื่อว่าต้นทุนของ Fault Negative และ Fault Positive แตกต่างกันมาก โมเดลเครือข่ายแบบเบย์สำหรับโรคเบาหวาน จะให้ค่า F1 สูงสุดในกรณีที่กำหนดเกณฑ์ (Threshold) เป็น 0.1 มีค่า คะแนน F1 คือ 0.2449 (โดยให้ค่าความถูกต้องของโมเดล (Accuracy) เป็น 80.50%) พบว่าการใช้โมเดลเครือข่ายแบบเบย์สำหรับโรคเบาหวานอาจจะไม่เหมาะสมนักเนื่องจากมีค่า คะแนน F1 ค่อนข้างต่ำ เมื่อพิจารณาผลการพยากรณ์ของโมเดลเครือข่ายแบบเบย์ สำหรับโรคความดันโลหิต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สูง พบว่ามีค่าคะแนน F1 สูงสุดเมื่อกำหนดระดับเกณฑ์ (Threshold) เป็น 0.2 ที่ทำให้มีค่า F1 เป็น 0.4412 (โดยให้ค่าความถูกต้องของโมเดล (Accuracy) เป็น 65.65%)

5.2 เหตุผลที่วิธีโมเดลเครือข่ายแบบเบย์เหมาะสมสำหรับการศึกษาด้านปัจจัยเสี่ยงการเกิดโรคไม่ติดต่อเรื้อรัง

การประยุกต์ใช้โมเดลเครือข่ายแบบเบย์ในการศึกษาในครั้งนี้มีข้อแตกต่างจากการใช้วิธีอื่นๆ โดยสรุปดังนี้

1. โมเดลสามารถใช้เพื่อการวิเคราะห์ได้หลายรูปแบบโดยเฉพาะอย่างยิ่งการพยากรณ์ และการวินิจฉัยโรค (Sebastiani et al., 2007) เช่นการพยากรณ์การเกิดโรค วิเคราะห์หาสาเหตุของการเกิดโรคจากสาเหตุความเสี่ยงต่างๆ หรือประมาณค่าผลกระทบของการเกิดโรคเมื่อกำหนดสถานการณ์ให้ความเสี่ยงหลายปัจจัยเกิดขึ้นพร้อมกัน ตามที่ได้นำเสนอไว้ข้างต้น

2. การอนุมานอาศัยทฤษฎีเบย์ (Bayes Theory) เป็นพื้นฐานในการพัฒนาโมเดลโดยใช้ความน่าจะเป็นแบบมีเงื่อนไขเพื่อหาความน่าจะเป็นร่วมจากตัวแปรที่เป็นสาเหตุเพื่อวิเคราะห์ว่าแต่ละปัจจัยส่งผลต่อกันและกันอย่างไร เมื่อเปรียบเทียบกับวิธี Neural Network ที่ไม่สามารถอธิบายกระบวนการอย่างชัดเจน ซึ่งเรียกว่า ‘Black-boxes model’ (Wagholikar et al., 2012)

3. โมเดลมีข้อจำกัดน้อยกว่าการใช้วิธีการทางสถิติทั่วไป เช่น การวิเคราะห์การถดถอย ที่ต้องมีเงื่อนไขการแจกแจงความเป็นปกติของข้อมูล หรือ การที่ตัวแปรอิสระห้ามมิให้มีความสัมพันธ์ระหว่างกันเอง (Multicollinearity) เป็นต้น

4. โครงสร้างโมเดลเครือข่ายแบบเบย์เป็นโครงสร้างแบบลำดับชั้น (Hierarchical Structure) แสดงความสัมพันธ์ลักษณะสาเหตุและผลลัพธ์ ซึ่งแตกต่างจากการใช้วิธี Logistic Regression ทั่วไป (Nguefack-Tsague, 2011) ที่นำเอาตัวแปรอิสระมาวิเคราะห์โดยจัดให้ตัวแปรทั้งหมดอยู่ในลำดับเดียวกัน ดังนั้นจึงไม่สามารถวิเคราะห์หาต้นกำเนิดของสาเหตุการณ์เกิดโรค (Root Cause) ได้อย่าชัดเจน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. โครงสร้างของโมเดลที่นำเสนอด้วยรูปกราฟนั้นง่ายที่จะเข้าใจ (Suvisaari et al., 2011) ดังนั้นโมเดลสามารถใช้ในการสื่อสารทั้งกลุ่มที่เป็นเชิงปฏิบัติถึงระดับนโยบายรวมทั้งบุคคลทั่วไปให้เข้าใจได้อย่างง่ายดาย

นอกจากนี้ยังพบว่าจากการศึกษาของ Atoui และคณะพบว่าโมเดลเครือข่ายแบบเบย์สามารถพยากรณ์การเกิดโรคหัวใจและหลอดเลือดจากฐานข้อมูลได้ดีกว่า Neural Network และ Logistic Regression

5.3 ข้อเสนอแนะ

โมเดลเครือข่ายแบบเบย์ที่สร้างขึ้นในงานวิจัยนี้ยังมีจุดที่ต้องการพัฒนาเพิ่มเติมในประเด็นต่างๆ ดังต่อไปนี้

1. จากผลการวิเคราะห์ของงานวิจัยนี้ระบุว่าสำหรับคนไทยที่มีประวัติครอบครัวโรคเบาหวานและเป็นโรคอ้วนควรถ่ายโอนความเสี่ยงเหล่านั้นไปยังตัวแทนอื่น ๆ ถึงแม้ว่ารัฐบาลไทยจะเสนอโครงการประกันสุขภาพ (เช่นหลักประกันสุขภาพถ้วนหน้าโครงการประกันสังคมและสวัสดิการด้านข้าราชการพลเรือน) อาจพิจารณาโอนส่วนที่เหลือให้กับบุคคลที่สาม นอกจากนี้บางคนที่ต้องการการดูแลเป็นพิเศษสามารถพิจารณาซื้อประกันสุขภาพจากบริษัทประกันสุขภาพในเชิงพาณิชย์

2. เนื่องจากตัวแบบที่สร้างขึ้นจากตัวแปร 7 ตัว สำหรับโมเดลเครือข่ายแบบเบย์ของการเกิดโรคเบาหวาน และ ตัวแปร 10 ตัว สำหรับโมเดลเครือข่ายแบบเบย์ของการเกิดโรคความดันโลหิตสูง ดังนั้นอาจมีตัวแปรบางตัวถูกตัดออกจากตัวแบบนี้ ตัวอย่างเช่น การสูบบุหรี่และการบริโภคเครื่องดื่มแอลกอฮอล์ เนื่องจากผลกระทบโดยตรงของตัวแปรเหล่านี้ไม่ชัดเจนจากเอกสาร (และจากความเห็นของผู้เชี่ยวชาญ) อย่างไรก็ตามโมเดลนี้สามารถปรับปรุงและปรับปรุงได้อย่างต่อเนื่อง

3. แนวทางการใช้โมเดลในการวิเคราะห์ความเสี่ยง สามารถนำไปประยุกต์ใช้กับโมเดลของทุกประเทศในการวิเคราะห์ความเสี่ยง

4. ถึงแม้ว่าผลการวิจัยนี้จะได้จากการสำรวจสุขภาพแห่งชาติในประเทศไทย ข้อมูลก็สามารถนำไปสู่การใช้ในประเทศอื่น ๆ ได้โดยการยืนยันแบบจำลองด้วยข้อมูลที่เกี่ยวข้องจากประเทศอื่น ๆ ในภูมิภาคเอเชียตะวันออกเฉียงใต้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. การพัฒนาโมเดลเครือข่ายแบบเบย์สามาถนำการสร้าโมเดลด้วยวิธีเรียนรู้จากข้อมูล เพื่อนำมาเปรียบเทียบกับโมเดลที่สร้างโดยอาศัยผู้เชี่ยวชาญที่ได้จากงานวิจัยนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6 สรุปผลผลิตที่ได้จากงานวิจัย

ผลผลิตจากงานวิจัยนี้ได้ดำเนินการนำส่วนหนึ่งของงานวิจัยไปนำเสนอในการประชุมวิชาการระดับนานาชาติ The IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) 2017

ระหว่าง rom **10-13 December, 2017 in Singapore ณSuntec Singapore Convention & Exhibition Centre** 1 Raffles Boulevard, Suntec City, Singapore 039593

โดยแสดงบทความที่นำเสนอในงานประชุมวิชาการดังกล่าวไว้ที่ภาคผนวก ก.

หมายเหตุ: อาจมีการเขียนบทความเพิ่มเติมเพื่อส่งวารสารในประเทศอีก 1 ฉบับ อยู่ระหว่างดำเนินการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

Aekplakorn, W., Abbott-Klafter, J., Premgamone, A., Dhanamun, B., Chaikittiporn, C., Chongsuvivatwong, V., ... Lim, S. S. (2007). Prevalence and management of diabetes and associated risk factors by regions of Thailand: Third National Health Examination Survey 2004. *Diabetes Care*, *30*(8), 2007–12. http://doi.org/10.2337/dc06-2319

Atoui, H., Fayn, J., Gueyffier, F., & Rubel, P. (2006). Cardiovascular risk stratification in decision support systems: A probabilistic approach. application to pHealth.

Burnside, E. S., Rubin, D. L., Fine, J. P., Shachter, R. D., Sisney, G. A., & Leung, W. K. (2006). Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience. *Radiology*, *240*(3), 666–73. http://doi.org/10.1148/radiol.2403051096

Danaei, G., Finucane, M. M., Lu, Y., Singh, G. M., Cowan, M. J., Paciorek, C. J., ... Ezzati, M. (2011). National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2·7 million participants. *Lancet*, *378*(9785), 31–40. http://doi.org/10.1016/S0140-6736(11)60679-X

Fellaji, S., Azmani, A., & Akharif, A. (2014). Bayesian approach for minimizing nephropathy risk for patients with type 2 diabetes. In *2014 9th International Conference on Intelligent Systems: Theories and Applications (SITA-14)* (pp. 1–4). IEEE. http://doi.org/10.1109/SITA.2014.6847311

Jensen, F. V, & Nielsen, T. D. (2007). *Bayesian networks and decision Graphs* (2nd ed.). New York: Springger.

Kjaerulff, U. B., & Madsen, A. L. (2008). *Bayesian networks and influence diagrams: a guide to construction and analysis*. New York: Springer.

Lauritzen, S. L., & Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, *50*(2), 157–224. Retrieved from http://www.jstor.org/stable/2345762

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Liu, K. F.-R., Lu, C.-F., Chen, C.-W., & Shen, Y.-S. (2011). Applying Bayesian belief networks to health risk assessment. *Stochastic Environmental Research and Risk Assessment*, *26*(3), 451–465. http://doi.org/10.1007/s00477-011-0470-z

Long, W. J., Naimi, S., Criscitiello, M. G., & Larsen, G. (1989). Differential diagnosis generation from a causal network with probabilities. In *Proceedings. Computers in Cardiology 1988* (pp. 185–188). IEEE Comput. Soc. Press. http://doi.org/10.1109/CIC.1988.72595

Marcot, B. (2012). Metrics for evaluating performance and uncertainty of Bayesian network models. *Ecological Modelling*, *230*(null), 50–62. http://doi.org/10.1016/j.ecolmodel.2012.01.013

Nguefack-Tsague, G. (2011). Using Bayesian Networks to Model Hierarchical Relationships in Epidemiological Studies. *Epidemiology and Health*, *33*, 1–8. Retrieved from https://www.scienceopen.com/document_file/a87cbc1d-5b9b-4c6d-81e8-3f18b0107104/PubMedCentral/a87cbc1d-5b9b-4c6d-81e8-3f18b0107104.pdf

Sebastiani, P., Nolan, V. G., Baldwin, C. T., Abad-Grau, M. M., Wang, L., Adewoye, A. H., … Steinberg, M. H. (2007). A network model to predict the risk of death in sickle cell disease. *Blood*, *110*(7), 2727–35. http://doi.org/10.1182/blood-2007-04-084921

Sodhi, M. S., & Tang, C. S. (2012). *Managing supply chain risk*. London: Springer.

Stephan, J., & Badr, Y. (2007). A quantitative and qualitative approach to manage risks in the supply chain operations reference. In *Second International Conference on Digital Information Management* (Vol. 1, pp. 410–417). Lyon: IEEE. http://doi.org/10.1109/ICDIM.2007.4444258

Suvisaari, J., Loo, B.-M., Saarni, S. E., Haukka, J., Perälä, J., Saarni, S. I., … Jula, A. (2011). Inflammation in psychotic disorders: a population-based study. *Psychiatry Research*, *189*(2), 305–11. http://doi.org/10.1016/j.psychres.2011.07.006

Twardy, C., Nicholson, A., & Korb, K. (2005). Knowledge engineering cardiovascular Bayesian networks from the literature. Retrieved from http://www.csse.monash.edu.au/publications/2005/tr-2005-170-full.pdf

Vose, D. (2008). *Risk analysis: a quantitative guide* (3rd ed.). Chichester: Wiley.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Wagholikar, K. B., Sundararajan, V., & Deshpande, A. W. (2012). Modeling paradigms for medical diagnostic decision support: a survey and future directions. *Journal of Medical Systems*, *36*(5), 3029–49. http://doi.org/10.1007/s10916-011-9780-4

Yang Guo, Guohua Bai, & Yan Hu. (2012). Using Bayes Network for Prediction of Type-2 diabetes. In *The 7th International Conference for Internet Technology and Secured Transactions (ICITST-2012)* (pp. 471–472).

สำนักงานสำรวจสุขภาพประชาชนไทย. (2552). *รายงานการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย: ครั้งที่ 4 พ.ศ. 2551-2552*. (เอกพลากรวิชัย, Ed.). กรุงเทพฯ: บริษัท เดอะ กราฟิโก ซิสเต็มส์ จำกัด. Retrieved from http://www.hiso.or.th/hiso/picture/reportHealth/report/report1.pdf

สำนักงานสำรวจสุขภาพประชาชนไทย. (2554). *การบริโภคอาหารของประชาชนไทย การสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างหายครั้งที่ 4 พ.ศ. 2551-2552*. (วิชัย เอกพลากร, Ed.). นนทบุรี.

สำนักโรคไม่ติดต่อ กรมควบคุมโรค. (2556). *รายงานประจำปี พ.ศ. 2557*. กรุงเทพฯ. Retrieved from https://www.google.co.th/?gws_rd=cr&ei=GOgoVMqTIoqRuATuzoGYBg#q=รายงานการเฝาระวังโรคไม่ติดต่อเรื้อรัง+พ.ศ.+2557

อมรา ทองหงษ์, กมลชนก เทพสิทธา, & ภาคภูมิ จงพิริยะอนันต์. (2556). รายงานการเฝ้าระวังโรคไม่ติดต่อเรื้อรัง ปี พ.ศ. 2554. รายงานการเฝ้าระวังทางระบาดวิทยาประจำสัปดาห์, *44*(10), 145–152. Retrieved from http://www.boe-wesr.net/index.php

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก. เอกสารหลักฐานอ้างอิงของผลผลิต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Applying Bayesian Network for Noncommunicable Diseases Risk Analysis: Implementing National Health Examination Survey in Thailand

K. Leerojanaprapa[1], W. Atthirawong[1], W. Aekplakorn[2] and K. Sirikasemsuk[3]
[1]Department of Statistics, King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, Thailand
[2]Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
[3]Department of Industrial Engineering, King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand
(kanogkan.le@kmitl.ac.th)

Abstract – **We propose using a Bayesian network to capture and understand the dependency risk factors affecting the prevalence of chronic diseases. By applying a Bayesian network model, we can visualize inter-dependencies between risks and their effects on the Noncommunicable disease (NCD) prevalence.**

By using a Bayesian network to model the prevalence of diabetes, we can define the top three risks as family history of diabetes, obesity, and age. Furthermore, the risk classification results can help to determine the managing strategy. For the Thai population, problems arising from family history of diabetes and obesity can be met by employing a transfer strategy. Age (especially ages of 35–59) and the risk incurred by low intake of fruits and vegetables should use a reduction or mitigation strategy. Finally, those at risk as a result of their area of residence (in urban areas) and socio-economic factors within the 4th quantile and low level of physical activity should apply a retain strategy.

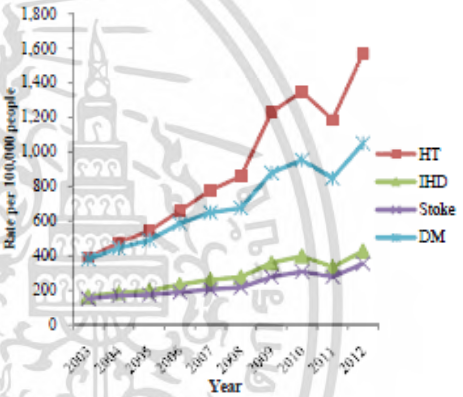
Keywords – **Bayesian network, noncommunicable diseases (NCDs), risk analysis**

I. INTRODUCTION

Noncommunicable, or chronic, diseases (NCDs) are diseases of long duration and generally slow progression. The World Health Organization defines the four main types of NCDs as cardiovascular diseases (like heart attacks and stroke), cancer, chronic respiratory diseases (such as chronic obstructed pulmonary disease and asthma) and diabetes. NCDs are increasingly prevalent and nowadays are the main cause of death, representing 63% of all annual deaths [1]. About 80% of all NCD deaths occur in low- and middle-income countries.

Thailand is one of the middle-income countries that is facing the rising rate of NCDs. Ischemic heart disease (IHD), hypertension (HT), diabetes (DM), and stroke are rising rapidly [2]. According to the *Annual Epidemiological Surveillance Report 2011* [2], the diseases with the highest rate per 100,000 people in 2012 was hypertension (HT), with 1,571 per 100,000, followed by DM (1,050.1 per 100,000), IHD (428 per 100,000), stroke and paroxysmal stroke (355 per 100,000); see Fig. 1.

Facing the burdens generated by the rise in NCDs has become one of the main medical and public health issues. In particular, the government is forced to focus on the main risks in order to reduce the incidence of NCDs, because patients need long-term treatment and the expenditure on NCD treatment keeps growing. The Ministry of Public Health is one of the main organizations to launch policies on reducing the incidence of NCDs, trying to promote the message that 'Prevention is better than cure' as prevention will help reduce the cost of screening, care and treatment. However, the incidence rates rise continuously.



Source: Annual Epidemiological Surveillance Report 2011 [2]

Fig. 1. Inpatient rate per 100,000 populations by major non-communicable diseases, 2003–2012

The prevention of NCDs is made difficult by the variety of causes of the diseases [3]. There are three main groups of risk factors: behavioral (e.g. smoking, physical inactivity, unhealthy diet, or alcohol drinking), metabolic (e.g. raised blood, high blood glucose levels, or obesity), and individual characteristics (e.g. age, sex, economic status). In addition, interactions between factors can be a catalyst for NCDs, such as if one is aware of being obese and yet does not exercise. The chances of someone in this situation becoming diabetic are greater than those who are obese and exercising. Therefore, in order to understand what are the major risks and how they interact, a risk model should be constructed to handle many risk factors and their interactions.

We propose employing a Bayesian network (BN) to capture causal and consequential factors of NCDs by explaining cause-effect relationships in a network model. Furthermore, a BN can also provide an analysis of the importance of risk factors in diagnosing the causes of

978-1-5386-0948-4/17/$31.00 ©2017 IEEE

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

NCDs and can then help in the drawing up of further plans to cope with significant risk factors.

Implementing a BN model for NCDs is not a new approach. BN methodology has been used to diagnose or predict the occurrence of various diseases such as breast cancer [4], anemia [5], cardiovascular disease [6], [7], and Type 2 diabetes [8]. But a limitation of the previous literature is that it does not include people with the disease who do not know that they have contracted it, or have not yet entered hospital. It cannot, therefore, be used to infer the prevalence of NCDs throughout the population. Furthermore, previous work does not include behavioral risks or other important characteristics that need to be considered.

Although some previous NCD studies employed experimental studies with patients, they focused on specific therapies. However, to show levels of NCD risk for the Thai population, data on the prevalence of the diseases should not just be collected from the general patient database of hospitals, as this excludes the characteristics of people who are not NCD patients or people who do not know their symptoms. For this reason, data from the *Thai National Health Examination Survey*, maintained by the Thai Public Health Survey Institute for Health Systems Research, are selected for this study.

Data obtained from the Health Survey consist of health risk factors, demographic characteristics, risk behaviors, and results from health examinations laboratory tests such as blood and urine tests. These data are from a large database that can be used to determine the biomarker, behavioral maker, and population indicator (characteristic marker).

Data from the Health Survey have never been used to model relationships between risk factors associated with NCDs, as the data presented in the report are focused on identifying single risk factors or identifying individual risk factors in a 2 or 3-way frequency classification scheme, as shown in [9], [10].

In this article, we discuss how a BN model can show the interaction between important risk factors to analyze the main causes of NCDs in the context of Thailand. We organize the article into four sections. Section II provides an overview of the methodology used to construct the BN for NCDs risk model. In Section III, we present a selection of the NCDs risk analysis. We summarize our conclusions and reflections in Section IV.

II. METHODOLOGY

A. Bayesian Network Modeling Concept

Fig. 2 shows the key concepts underpinning our approach to model NCD risks. The diagram represents the cause-effect relations between risk factors. The "effect" corresponds to the NCD. The "cause" includes three main root sources: metabolic risk factors, behavioral risk factors, and demographic risk factors.

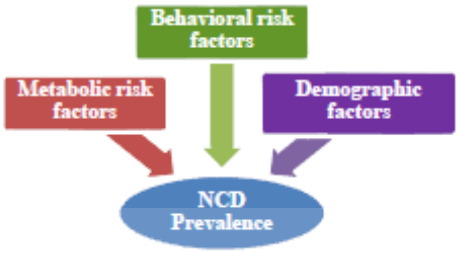


Fig. 2. NCD risk modeling concept

B. Bayesian Network Theory and Example

We adopt BN theory because this can help to represent uncertain relationships between risk events; the uncertainty regarding the occurrence of a risk event can be quantified as a probability; and analysis grounded in the theory of Bayesian methods can be conducted [11].

A BN can be constructed to represent different types of relationship; causal relationships will be the focus of this study. An arrow links from a direct cause to its effect. For example, if Family History (C_1) and Obesity (C_2) are causes of Diabetes, the relation within a BN can be represented as in Fig. 3.

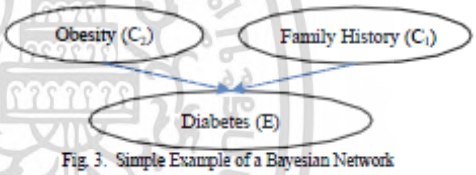


Fig. 3. Simple Example of a Bayesian Network

Since the relation can represent uncertain effects, a natural way of specifying the probabilistic relations among those three variables can be quantified by probability and conditional probability. According to the example above, some obese people who have some of their family suffering from diabetes themselves have diabetes, but some do not. Therefore, the uncertain nature of the causal relation is shown by conditional probability, for example as shown in Table 1.

TABLE I
PROBABILITY AND CONDITIONAL PROBABILITY OF THE VARIABLES IN BAYESIAN NETWORK EXAMPLE

$P(C_2)$	
No	0.65
Yes	0.35

$P(C_1)$	
No	0.75
Yes	0.25

$P(E\|C_1,C_2)$	C_1 = No		C_1 = Yes	
	C_2 = No	C_2 = Yes	C_2 = No	C_2 = Yes
No	0.9	0.3	0.2	0.1
Yes	0.1	0.7	0.8	0.9

C. Inference in Bayesian Networks

BN theory is grounded in the conditional probability distribution and conforms to Bayes' rule [12], as in (1).

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \quad (1)$$

where

$P(A)$ is the prior distribution
$P(B|A)$ is the conditional probability distribution
$P(A|B)$ is the posterior probability distribution

In the example above, if we assume that C_1 (Family History) and C_2 (Obesity) are the only two causes of E (diabetes), then:

$$P(C|E) = \frac{\sum_{C} P(E|C_i, C_j) * P(C_i) * P(C_j)}{P(E)} \quad (2)$$

D. Normalised Likelihood

Normalised likelihood (NL) is a sensitivity analysis method [13]. Diagnostic reasoning from the top event (Y) to individual cause variables (ε) is implemented by calculating marginal posterior probabilities of the variable, ε. The ratio of the adjusted probability of the given variable and current probability is defined in the NL equation to help prioritize significant risks [14]:

$$NL = \frac{P(\varepsilon|T)}{P(\varepsilon)} \quad (3)$$

where

$P(\varepsilon)$ is the current probability (without observing the top event).
$P(\varepsilon|T)$ is the adjusted probability (given that a focus variable has been observed.

E. Risk Prioritization

This article uses the results of BN inferential analysis by normalized likelihood to compare each factor affecting the disease. Based on the difference between the adjusted probability and the current probability of individual variables as expressing by NL, it can be used to prioritise all risk factors in the model.

F. Risk Classification

The classification of risk factors in general research is often displayed in the form of a risk matrix. A risk matrix can categorize risks, since a risk component consists of the probability (Probability) and the impact of each factor (Risk) to classify the risk factors. This can support decision making when it comes to selecting proper approaches to manage risks. General risk mitigating actions have been clearly explained in [15], especially in terms of implementing different risk mitigating strategies in different situations.

1. *Risk avoidance* is suitable for risks that can happen frequently and, when they occur, can generate high impacts on the organisation.
2. *Risk transfer* aims to shift bad consequences to others by purchasing insurance or outsourcing.
3. *Risk reduction/mitigation* is suitable for any level of risk that is not severe (high probability and high impact) by trading off between benefits and costs.
4. *Risk retention* is suitable for risks that are not significant because they have both low likelihood and impact, compared with the cost of control.

The approach of using BN analysis to categorise risks is also presented by [14], representing impact by percentage probability increase and probability by current probability of occurrence for individual variables, as shown in Fig. 4.

Impact (% of increase)		
High	Transfer	Avoidance
Low	Retain/ Absorption	Reduction/ Mitigation
	Low	High

Probability (Current probability)

Fig. 4. Risk Matrix

G. Key Stages of Bayesian Network Modeling

Our process to build a BN for NCDs, summarized in Table II, is informed by adjustments from general approaches for BN modeling [14], [16], [17]. There are seven stages in the process..

Stage 1: Identify scope of the model by interview – In the case study, a professor in preventive medicine who was a member in a health survey team was interviewed. We defined the scope as Type 2 diabetes for the Thai population of 15 years old and above.

Stage 2: Structure the provisional BN risk model – Two professors in family medicine participated in this research: they were interviewed and the results used to structure the BN model.

Stage 3: Refine the BN NCD risk structure – A workshop with all experts was held, which allowed a review of the BN structure and the definition of model variables.

Stage 4: Quantify the BN risk model from the Health Survey data – Data in input format were extracted from the Thai National Health Examination Survey.

Stage 5: Input probability into the model and verify the model – Manually input data into the model, structured in GeNIe [18], and recheck input data to verify the model.

Stage 6: Use BN to support NCD risk analysis – Use GeNIe to perform diagnostic propagation to calculate the adjusted probability and then calculate NL.

Stage 7: Validate the BN model behavior – Face validity was defined in the group meeting with the experts. Scenario analysis can be used to explore meaningful cases defined by experts to investigate the behavior of the possible risks.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

TABLE II
PROCESS TO DEVELOP BAYESIAN NETWORK NCD RISK MODEL

Stage	Activity	Method
1	Identify scope of the model	Interview
2	Structure the provisional BN NCD risk model	Align with BN formalism
3	Refine the BN NCD risk structure	Workshop
4	Quantify the BN risk model from the Health survey data	Interview
5	Input probability into the model and verify the model	Verification
6	Use BN to support NCD risk analysis	Analysis with software tool
7	Validate the BN model behavior	Workshop

III. RESULTS FROM NCD RISK MODEL

The outcomes from this research can be divided into two main parts: diabetes BN model structure and results of risk analysis.

A. *Diabetes BN Model Structure*

Fig. 5 shows the relationship between selected risk factors to the prevalence of diabetes. The overall model contains eight nodes, representing risk events, and eleven arcs, capturing relationships between these events.

The top event is defined as the prevalence of diabetes and this event might be due to multiple causes. The direct causal relations were defined by the experts. They defined the eight relationships as age, DM family history, area of residence, socio-economic factors, fruit and vegetable consumption, obesity, and physical activities. The defined states of particular variables in the model are defined and suggested by the experts and we also concern the available information from the National Health Examination Survey data. 'Socio-economic factors' is classified into five states by the quantile of wealth index score. 'Fruit and vegetable consumption' is classified by number of servings per day. 'Obesity' is classified by body mass index (BMI) and waist circumference. Physical activity is defined by types of physical activity, frequency, and duration of activity in a regular week.

Fig. 5. Diabetes BN model

B. *Result of Risk Analysis*

Risk Prioritization

All defined risk factors in the BN model with representing state are prioritized by considering the NL or percentage increase, as shown in Table III.

TABLE III
OUTPUT OF RISK PRIORITISATION

Risk Variable	State	Current	Adjusted	NL	% of Increase	Rank
Age	35-59 year	0.6664	0.8942	1.3419	34.1851	3
DM Family History	Yes	0.2524	0.5220	2.0682	106.8219	1
Area of Residence	Urban	0.3016	0.3025	1.0028	0.2824	7
Socio-economic	4th Quantile	0.2000	0.2007	1.0034	0.3414	6
Fruit and Vegetable consumption	Less*	0.8107	0.8143	1.0044	0.4407	5
Obesity	Over BMI**	0.3547	0.5941	1.6749	67.4931	2
Physical activity	Low***	0.1651	0.1743	1.0556	5.5588	4

* < 5 servings/day

** BMI ≥ 25 or Waist circumference ≥ 90 cm (male) / ≥ 80 cm (female)

*** a. Heavy activity ≥ 3 days/week and time ≥ 20 minutes a day, b. Moderate activity ≥ 5 days and time ≥ 30 minutes a day

For example, the chances of a person having a diabetes DM family history increase from 0.2524 to 0.5220 if a person has diabetes. As a result, the ratio increased to 2.0682 times, representing an increase of 106.8219%. This percentage increase can be used for risk prioritization.

From Table III it is clear that people who have DM family history have the highest percentage increase. Obesity (over BMI), age (35–59 year), physical activity (low), fruit and vegetable consumption (less), and area of residence (in urban areas) are ranked respectively.

Risk Classification

The results from Table II are employed to classify risk factors in order to define a suitable strategy for managing them (see Fig. 6).

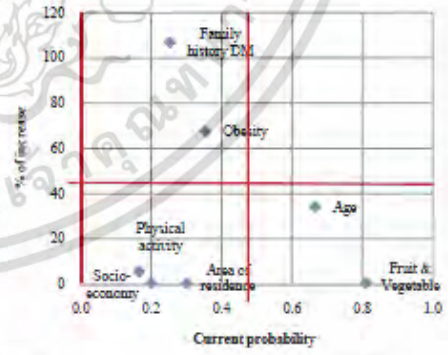


Fig. 6. Percentage increase and current probability of risk factors

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Diabetes risk management guidelines can be determined from the risk matrix which is divided by level of probability and impact. The graph can be divided into quadrants according to the high and low divisions of both, using a level of 0.5, as shown in Fig. 6. The four quadrants are defined as High probability and High impact, Low probability and High impact, High probability and Low impact and Low probability and Low impact.

For the Thai population, DM family history and obesity are found to have Low probability and High impact so we suggest employing a transfer strategy to cope with them, such as the introduction of health insurance that can cover the treatment of diabetes.

Age, especially between 35 and 59, and the low intake of fruits and vegetables are two risk factors within the High probability and Low impact group; therefore, risk reduction or mitigation can be the suggested strategy.

The area of residence and socio-economic factors lie within the 4th quartile, and are variables that represent the group of middle to high income. This does not represent the majority of Thai people, but the expert defined them as the risk group. In addition, low physical activity shows higher impact than the other two variables but they may be classified in the same group of Low probability and Low impact group. Thus, retain/absorption is the suitable strategy to manage them. It may not be necessary for the decision makers to focus on those factors.

There is no risk variable within the High probability and High impact quadrant.

IV. REFLECTIONS AND CONCLUSIONS

The presenting result is only one aspect to show the usefulness of applying BN in the context of NCDs. Our analysis indicates that for Thai population who has a DM family history and is obese should transfer those risks to other agents. Although, Thai government provides health insurance scheme (e.g. universal health coverage, social security scheme, civil servant medical benefit scheme), they may consider to transfer fractions of it to other third parties. Furthermore, some of them who want the extra care can consider buying health insurance from commercial insurers. For those with low intake of fruits and vegetables, the policy maker should promote them to eat more. According to the results, the impact and probability of obesity is in the middle of the matrix, Fig.6; therefore some decision makers can consider avoidance or mitigation strategy to cope with this risk factor instead of using risk transfer.

Some variables are omitted from this model, for example smoking and alcohol consumption, since the direct effects of these variables are unclear from the literature (and from the expert's opinion). However, the model can be continually improved and updated.

The results of this paper can help decision makers in health policy to prioritize and classify DM risk. Although the results from this model are generated from the National Health Examination Survey in Thailand, they can be generalized to other countries by confirming the model with relevant data from other countries in the Southeast Asia region.

ACKNOWLEDGMENTS

We would like to thank Ramathibodi Hospital staff who participated in the model building. We acknowledge the sponsorship from King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand, for their research grant.

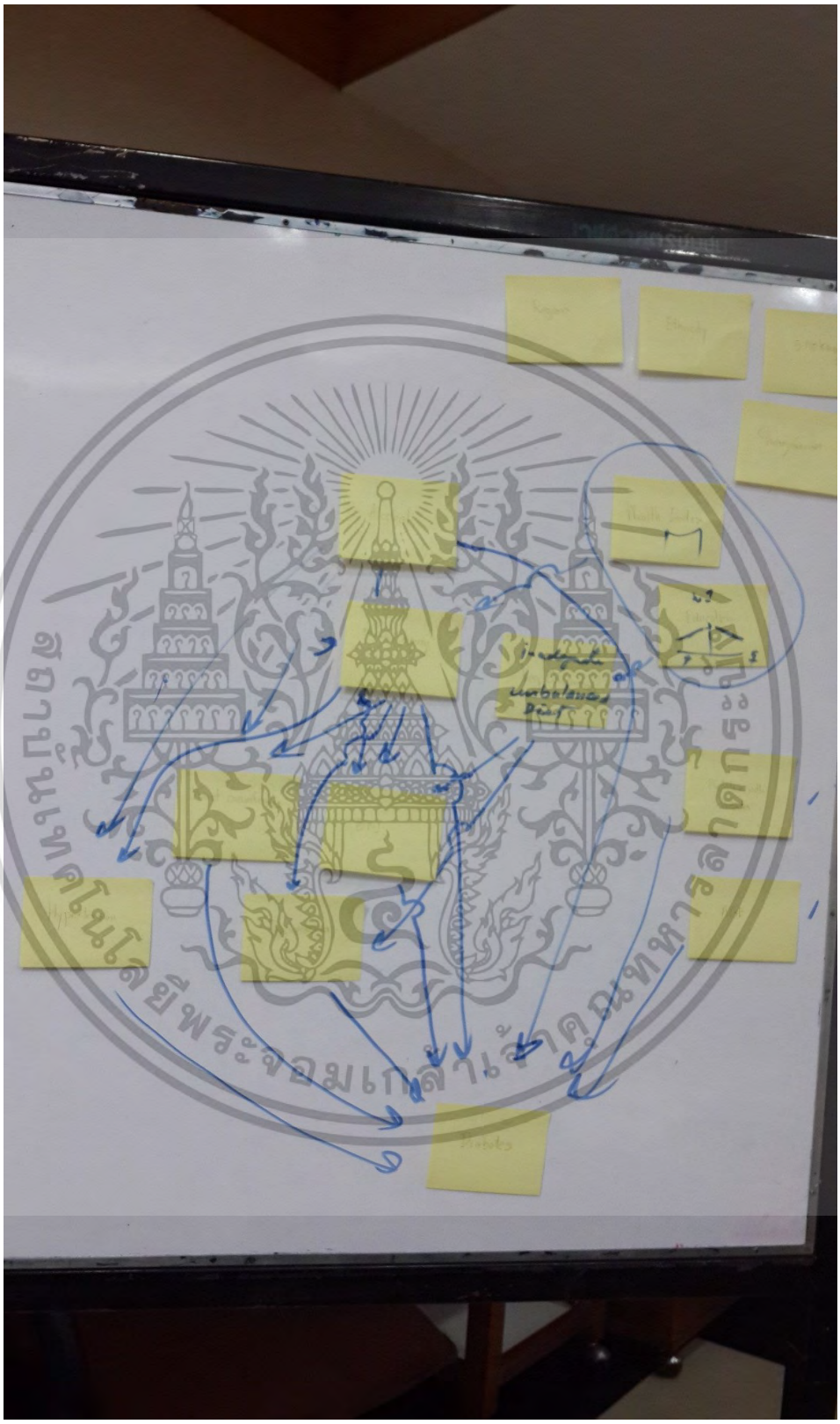
REFERENCES

[1] World Health Organisation (WHO), "WHO | 10 facts on noncommunicable diseases," WHO, 2014. [Online]. Available: http://www.who.int/features/factfiles/noncommunicable_diseases/en/.
[2] Bureau of Epidemiology, Ministry of Public Health, "Weekly Epidemiological Surveillance Report, Thailand," Annu. Epidemiol. Surveill. Rep. 2011, vol. 44, no. 10, pp. 146–159, 2013.
[3] World Health Organisation (WHO), "WHO | Noncommunicable diseases," WHO, 2017. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs355/en/.
[4] E. S. Burnside, D. L. Rubin, J. P. Fine, R. D. Shachter, G. A. Sisney, and W. K. Leung, "Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience.," Radiology, vol. 240, no. 3, pp. 666–73, Sep. 2006.
[5] P. Sebastiani, V. G. Nolan, C. T. Baldwin, M. M. Abad-Grau, L. Wang, A. H. Adewoye, L. C. McMahon, L. A. Farrer, J. G. Taylor, G. J. Kato, M. T. Gladwin, and M. H. Steinberg, "A network model to predict the risk of death in sickle cell disease.," Blood, vol. 110, no. 7, pp. 2727–35, Oct. 2007.
[6] H. Atoui, J. Fayn, F. Gueyffier, and P. Rubel, "Cardiovascular risk stratification in decision support systems: A probabilistic approach. application to pHealth," pp. 281–284, 2006.
[7] C. Twardy, A. Nicholson, and K. Korb, "Knowledge engineering cardiovascular Bayesian networks from the literature," 2005.
[8] Yang Guo, Guohua Bai, and Yan Hu, "Using Bayes Network for Prediction of Type-2 diabetes," pp. 471–472, 2012.
[9] National Health Examination Survey Office, "Thai National Health Examination Survey IV," Bangkok, 2009.
[10] W. Aekplakorn, J. Abbott-Klafter, A. Premgamone, B. Dhananun, C. Chaikittiporn, V. Chongsuvivatwong, T. Suwanprapisa, W. Chaipornsupaisan, S. Tiptaradol, and S. S. Lim, "Prevalence and management of diabetes and associated risk factors by regions of Thailand: Third National Health Examination Survey 2004.," Diabetes Care, vol. 30, no. 8, pp. 2007–12, Aug. 2007.
[11] K. Leerojanaprapa, R. Van Der Meer, and L. Walls, "Modeling supply risk using belief networks: A process with application to the distribution of medicine," in IEEE International Conference on Industrial Engineering and Engineering Management, 2014.
[12] F. V Jensen and T. D. Nielsen, Bayesian networks and decision Graphs, 2nd ed. New York: Springger, 2007.
[13] U. B. Kjaerulff and A. L. Madsen, Bayesian networks and influence diagrams: a guide to construction and analysis. New York: Springer, 2008.
[14] K. Leerojanaprapa, "A Bayesian belief network modelling process for systemic supply chain risk," University of Strathclyde., 2014.
[15] D. Vose, Risk analysis: a quantitative guide, 3rd ed. Chichester: Wiley, 2008.
[16] K. Leerojanaprapa, R. Van Der Meer, and L. Walls, "Modeling supply risk using Belief networks: A process with application to the distribution of medicine," in International Conference on Industrial Engineering and Engineering Management, 2013, p. 37.
[17] K. Leerojanaprapa, L. Walls, and R. Van der Meer, "Modelling and managing systemic risks in supply chains," in 18th International Annual EurOMA Conference Exploring interfaces, 2011, p. 112.
[18] The Decision Systems Laboratory, "GeNIe and SMILE." 2013. [Online]. Available: http://genie.sis.pitt.edu/about.html.

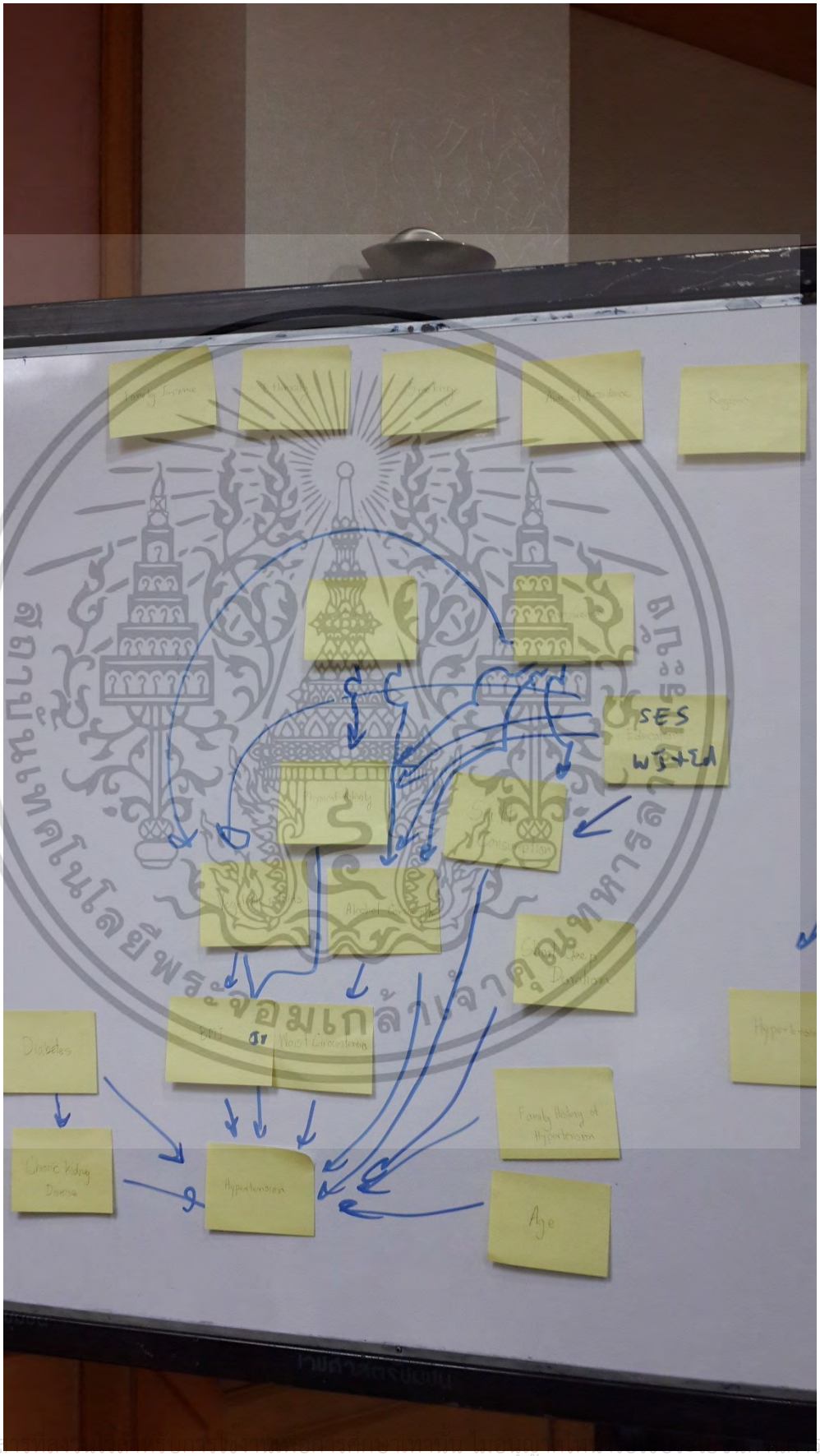
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข. โมเดลที่ได้จากการทำ Workshop

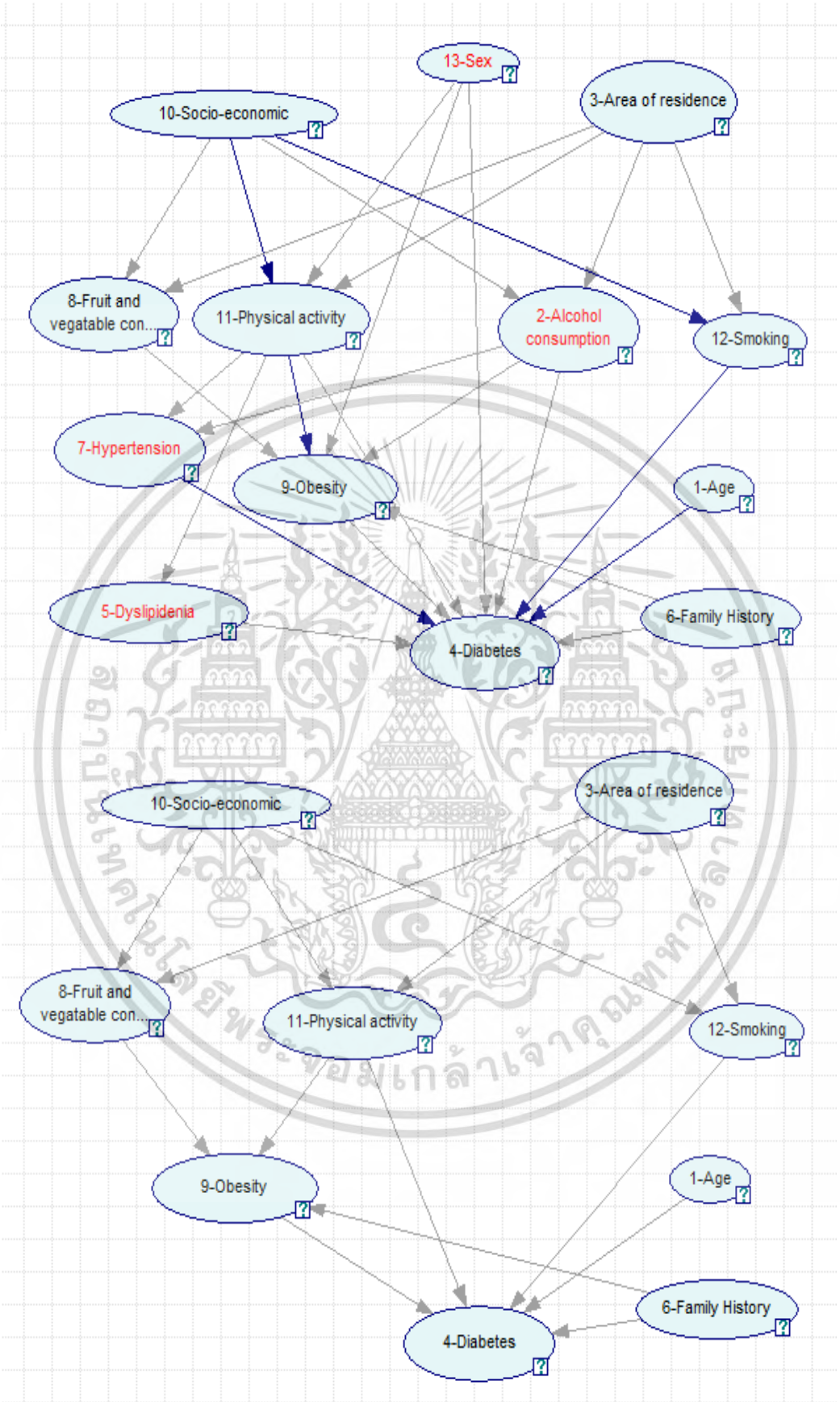
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



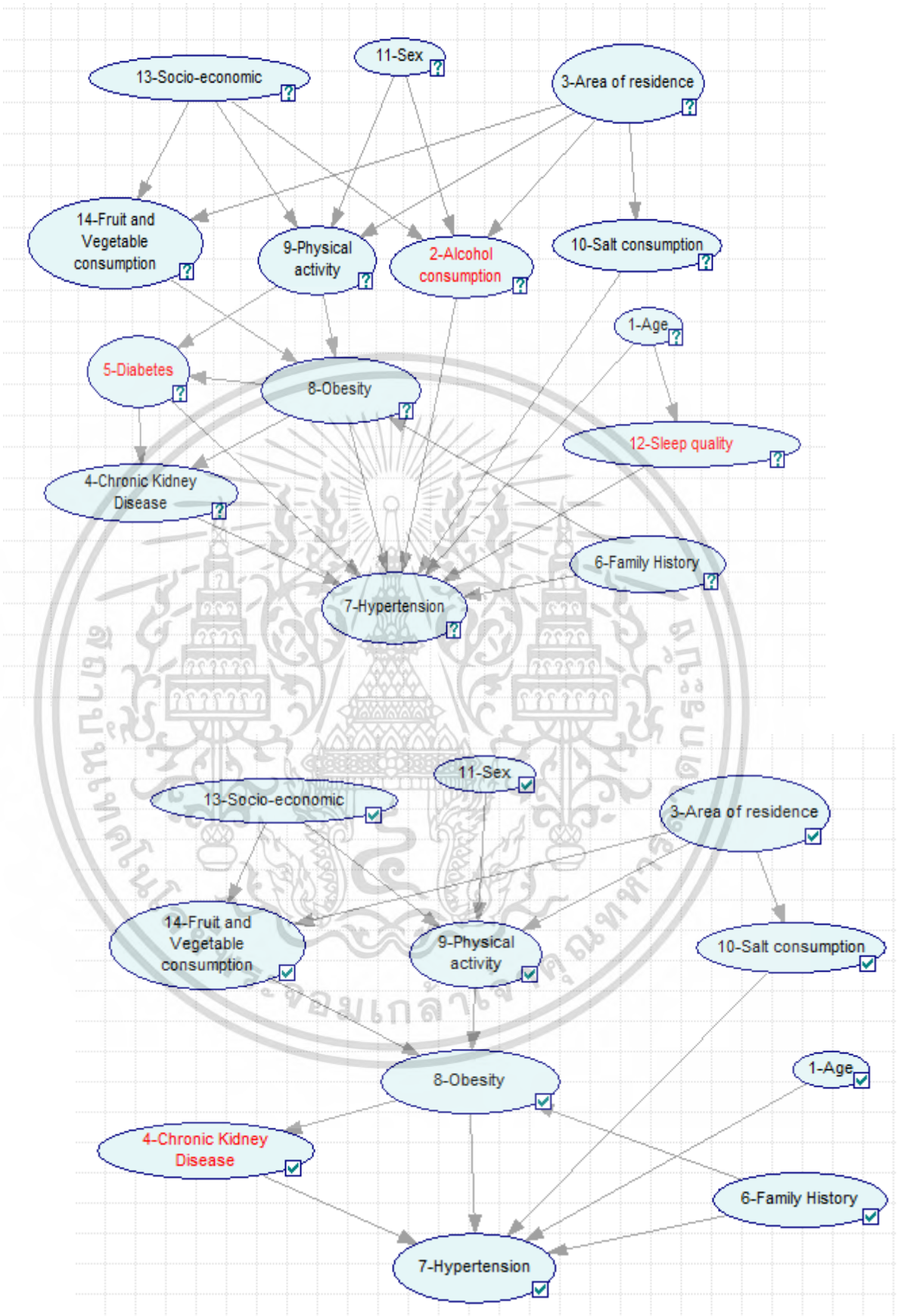
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



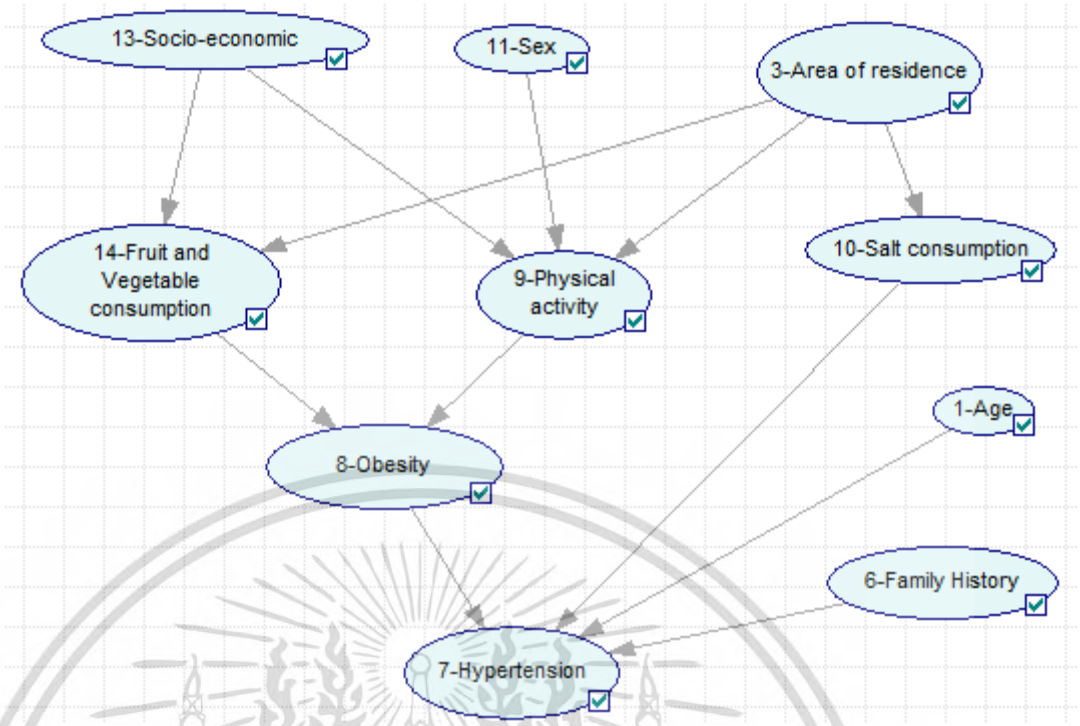
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ค. สรุปการใช้จ่ายเงิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สัญญาเลขที่ KREF015805

โครงการ การประยุกต์ใช้โมเดลเครือข่ายแบบเบย์เพื่อวิเคราะห์ความเสี่ยงการเป็นโรคไม่ติดต่อจากการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย

Applying Bayesian Network for Chronic Diseases Risk Analysis from National Health Examination Survey in Thailand

รายงานสรุปการเงินรอบ.....เดือน

ชื่อหัวหน้าโครงการวิจัย นางสาวกนกกรรณ์ ลี้โรจนาประภา

รายงานในช่วงตั้งแต่วันที่ 1 ตุลาคม 2558 ถึงวันที่ 30 กันยายน 2561

สรุปงบประมาณค่าใช้จ่ายที่ใช้นับตั้งแต่เริ่มทำการวิจัยถึงปัจจุบัน

หมวดค่าใช้จ่าย	งบประมาณรวมทั้งโครงการ	ค่าใช้จ่ายจากรายงานครั้งก่อน	ค่าใช้จ่ายงวดปัจจุบัน	รวมค่าใช้จ่ายสะสมถึงปัจจุบัน	คงเหลือ (หรือเกิน)
งบบุคลากร :ค่าจ้างชั่วคราว					
งบดำเนินงาน					
ค่าตอบแทน	25,000		25,000	25,000	0
ค่าใช้สอย	66,500		70,000	70,000	0
ค่าวัสดุ	13,300		14,000	14,000	0
ค่าสาธารณูปโภค	-				
งบลงทุน: ค่าครุภัณฑ์	-				
รวม	104,800		109,000	109,000	4,200

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จำนวนเงินที่ได้รับและจำนวนเงินที่ใช้จ่าย

งวดเงินที่ได้รับ	จำนวนเงินที่ได้รับ (บาท)	เมื่อ (ระบุวัน เดือน ปี)
งวดที่ 1	104,800	27 ตุลาคม 2558
งวดที่ 2	-	
ดอกเบี้ย ครั้งที่ 1	94.75	31 ธันวาคม 2558
ดอกเบี้ย ครั้งที่ 2	260.80	30 มิถุนายน 2559
ดอกเบี้ย ครั้งที่ 3	264.33	31 ธันวาคม 2559
ดอกเบี้ย ครั้งที่ 4	261.38	30 มิถุนายน 2560
ดอกเบี้ย ครั้งที่ 5	266.37	31 ธันวาคม 2560
ดอกเบี้ย ครั้งที่ 6	262.39	30 มิถุนายน 2561
รวม	106,210.32	❶

งวดที่	**จำนวนเงินที่ใช้จ่าย (บาท)**	
งวดที่ 1	104,800	
งวดที่ 2	-	
ฯลฯ		
รวม	104,800	❷

จำนวนเงินคงเหลือ ❶ - ❷ 1,410.32 **บาท**

ลงนามหัวหน้าโครงการวิจัยผู้รับทุน

ลงนามเจ้าหน้าที่การเงินโครงการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัตินักวิจัย

ประวัติส่วนตัว

ชื่อ-สกุล นางสาว กนกกรรณ์ ลี้โรจนาประภา

หน้าที่การงานปัจจุบัน ผู้ช่วยศาสตราจารย์

ประวัติการศึกษา

ชื่อย่อปริญญา	สาขา	สถาบันที่จบ	ปีที่จบ
วท.บ. B.Sc. (Statistics)	สถิติ	มหาวิทยาลัยธรรมศาสตร์	2542
สต.ม. M.Sc. (Statistics)	สถิติ	จุฬาลงกรณ์มหาวิทยาลัย	2545
PhD.	Management Science/ Operations Research	University of Strathclyde, UK	2557 (2014)

ประสบการณ์วิจัยหรือสาขาที่ชำนาญ

- ความเสี่ยงในห่วงโซ่อุปทาน (Supply Chain Risk)
- โมเดลเครือข่ายแบบเบย์ (Bayesian Network)
- สถิติ (Statistics)
- ความน่าจะเป็น (Probability)
- การสำรวจตัวอย่าง (Sampling Survey)
- สถิติวิเคราะห์ธุรกิจ

รางวัลด้านวิชาการ/ด้านวิจัย/งานสร้างสรรค์ (ด้านศิลปะ หรืออื่นๆ) ที่ได้รับ

ปี พ.ศ.	ชื่อรางวัล	สถาบันที่ให้
2559	รางวัลบทความดีมาก ในงานประชุมวิชาการด้านการวิจัยดำเนินงานแห่งชาติ ประจำปี 2559	สถาบันบัณฑิตพัฒนบริหารศาสตร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทุนการศึกษาและทุนวิจัยที่เคยได้รับ

ปี พ.ศ.	ทุนการศึกษาและทุนวิจัย	สถาบันที่ให้
2543	ทุนพัฒนาอาจารย์ ศึกษาระดับปริญญาโทในประเทศ	คณะกรรมการอุดมศึกษา
2548	ได้รับทุนส่งเสริมนักวิจัย เรื่อง “ปัจจัยที่มีผลต่อการสำเร็จการศึกษาและผลสัมฤทธิ์ทางการเรียนของนักศึกษาสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง” ปี 2548	คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
2549	ทุนพัฒนาอาจารย์ ศึกษาระดับปริญญาเอกต่างประเทศ	กระทรวงวิทยาศาสตร์และเทคโนโลยี
2558	ทุนวิจัยโครงการ “การเตรียมความพร้อมด้านกำลังคนเพื่อสร้างโซ่คุณค่าอุตสาหกรรมสิ่งทอและเครื่องนุ่งห่มไทย ในการเข้าสู่ประชาคมเศรษฐกิจอาเซียน	ทุนอุดหนุนการวิจัยจากสำนักงานกองทุนสนับสนุนการวิจัย (สกว.) ประจำปี
2558	ทุนวิจัยโครงการ “การวิเคราะห์ปัจจัยที่มีผลต่อการตัดสินใจเปลี่ยนแปลงการขนส่งเพื่อการส่งออกของอุตสาหกรรมสำคัญจากท่าเรือน้ำลึกแหลมฉบัง ท่าเรือกรุงเทพ เป็นท่าเรือน้ำลึกทวาย” ภายใต้แผนโครงการ “การศึกษาการจัดกลุ่มอุตสาหกรรม และผลกระทบต่อระบบโลจิสติกส์เพื่อการส่งออกบนแนวเส้นทางเชื่อมโยงทวาย – ระยอง”	สำนักงานกองทุนสนับสนุนการวิจัย (สกว.) ประจำปี 2558
2559	ทุนวิจัยโครงการ “การประยุกต์ใช้โมเดลเครือข่ายแบบเบย์เพื่อวิเคราะห์ความเสี่ยงการเป็นโรคไม่ติดต่อจากการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย”	ทุนพัฒนานักวิจัยใหม่ กองทุนวิจัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ประจำปี 2559 (สัญญาเลขที่ KREF015805)
2559	“การประยุกต์ใช้โมเดลเครือข่ายแบบเบย์ด้วยวิธีการเรียนรู้ของเครื่องเพื่อวิเคราะห์ความเสี่ยงการเป็นโรคไม่ติดต่อจากข้อมูลการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย”	คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
2560	การเปรียบเทียบการจัดกลุ่มอุตสาหกรรมในปัจจุบันกับคลัสเตอร์อุตสาหกรรมตามนโยบายของประเทศไทย”	คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลงานวิจัย/งานสร้างสรรค์ —

ผลงานวิจัย/งานสร้างสรรค์ที่ตีพิมพ์เผยแพร่ (ระดับชาติและนานาชาติ) —

วลัยลักษณ์ อัตธีรวงศ์, & กนกกรรณ์ ลี้โรจนาประภา. (2559a). การวัดประสิทธิภาพโรงงานอุตสาหกรรมที่เกี่ยวกับการเตรียมเส้นใยสิ่งทอด้วยวิธีโอบล้อมข้อมูล. In การประชุมด้านการวิจัยดำเนินงานแห่งชาติ ประจำปี 2559 (pp. 284–289). กรุงเทพฯ.

วิชิต หล่อจิระชุณกุล, วลัยลักษณ์ อัตธีรวงศ์, & กนกกรรณ์ ลี้โรจนาประภา. (2559b). ศักยภาพด้านโลจิสติกส์และโซ่อุปทานของผู้ประกอบการผ้าไหม. In การประชุมสัมนาเชิงวิชาการ ด้านการจัดการโลจิสติกส์และโซ่อุปทาน ครั้งที่ 16 (p. 36). กรุงเทพฯ.

วลัยลักษณ์ อัตธีรวงศ์, & กนกกรรณ์ ลี้โรจนาประภา. (2559c). การวัดประสิทธิภาพและการปรับปรุงประสิทธิภาพโรงงานอุตสาหกรรมการเตรียมเส้นใยสิ่งทอด้วยวิธีโอบล้อมข้อมูล. วารสารไทยการวิจัยดำเนินงาน, 4(1), 38–48.

วลัยลักษณ์ อัตธีรวงศ์, วิชิต หล่อจิระชุณกุล, & กนกกรรณ์ ลี้โรจนาประภา. (2560). การเปรียบเทียบศักยภาพด้านโลจิสติกส์และโซ่อุปทานระหว่างผู้ประกอบการผ้าไหมที่มีการรวมกลุ่มและไม่มีการรวมกลุ่ม. วารสารมหาวิทยาลัยนเรศวร, 10(3), 81–92.

วิชิต หล่อจิระชุณกุล, วลัยลักษณ์ อัตธีรวงศ์, & กนกกรรณ์ ลี้โรจนาประภา. (2561). การวิเคราะห์สภาวะแวดล้อมและเมททริกซ์ทาวช์เพื่อพัฒนายุทธวิธีการเพิ่มศักยภาพโซ่อุปทานผ้าไหมไทย-ลาว. วารสารการจัดการ, (3).

การเสนอผลงานวิชาการ

Leerojanaprapa, K. (2006). Predicting Bachelor's Degree Graduation Case Study on Faculty of Science, KMITL, Thailand. In Proceedings of the 11th Annual Conference of Asia Pacific Decision Sciences Institute (pp. 687–690). Retrieved from http://iceb.nccu.edu.tw/proceedings/APDSI/2006/687-690.pdf

Leerojanaprapa, K., Walls, L., & Van der Meer, R. (2011). Modelling and managing systemic risks in supply chains. In Eighteenth International Annual EurOMA Conference Exploring interfaces (p. 112). Cambridge: University of Cambridge Institute for Manufacturing. Retrieved from http://www.euroma2011.org/documents/abstracts_2011.pdf

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Leerojanaprapa, K. (2012). Comparing methods for structuring Bayesian belief network models for supply chain risk. In Informs International Beijing 2012 (p. 102). Beijing: INFORMS.

Leerojanaprapa, K., Van Der Meer, R., & Walls, L. (2013). Modeling supply risk using Belief networks: A process with application to the distribution of medicine. In International Conference on Industrial Engineering and Engineering Management (p. 37). Bangkok: IEEE.

Leerojanaprapa, K. (2014). A Bayesian belief network modelling process for systemic supply chain risk. University of Strathclyde.

Leerojanaprapa, K., & Atthirawong, W. (2015a). Management Practices of Thai Silk Product. In M. Gen, K. J. Kim, X. Huang, & Y. Hiroshi (Eds.), Industrial Engineering, Management Science and Applications 2015 (pp. 511–519). Tokyo: Springer Berlin Heidelberg. Retrieved from http://www.springer.com/us/book/9783662471999

Walls, L., Leerojanaprapa, K., & Der, R. van. (2015b). A Bayesian Network Model with Epistemic Uncertainty: Analysis of a Medicine Supply Chain Risk. In 27th European Conference on Operational Research Conference Handbook (p. 111). Retrieved from http://euro2015.euro-online.org/conference-homepage.html

Atthirawong, W., Sirovetnukul, R., Leerojanaprapa, K., Panprung, W., & Ruangteprat, T. (2016). Imbalance between Demand and Supply of Manpower for Textile Industry in Thailand. In the 7th International conference on operation and supply chain management (pp. 680–690). Phuket, Thailand.

Leerojanaprapa, K., & Bhundarak, K. (2017a). Resource Factors Affecting Labour Demand for Textile and Garment Industry in Thailand. Journal of Industrial Engineering and Management Science, 1, 263–288. http://doi.org/10.13052/jiems2446-1822.2017.013

Leerojanaprapa, K., Sirikasemsuk, K., & Bhundarak, K. (2017b). Factors Associated with Thai Exporter's Interest In Using New Dawei Deep Seaport. In 31th European Conference on Modelling and Simulation (pp. 127–134). Budapest, Hungary.

Sirikasemsuk, K., & Leerojanaprapa, K. (2017c). One missing value problem in Latin square design of any order: Exact analysis of variance. Cogent Engineering, 4(1). http://doi.org/10.1080/23311916.2017.1411222

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Sirikasemsuk, K., & Leerojanaprapa, K. (2017d). Regression sum of squares of symmetric balanced incomplete block design consisting of another one missing observation by accident. In 2017 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS) (pp. 233–238). IEEE. http://doi.org/10.1109/ICIIBMS.2017.8279744

Leerojanaprapa, K., Atthirawong, W., Aekplakorn, W., & Sirikasemsuk, K. (2018a). Applying Bayesian network for noncommunicable diseases risk analysis: Implementing national health examination survey in Thailand. In IEEE International Conference on Industrial Engineering and Engineering Management (Vol. 2017–Decem). http://doi.org/10.1109/IEEM.2017.8290023

Sirikasemsuk, K., & Leerojanaprapa, K. (2018b). Regression sum of squares of symmetric balanced incomplete block design consisting of another one missing observation by accident. In ICIIBMS 2017 - 2nd International Conference on Intelligent Informatics and Biomedical Sciences (Vol. 2018–Janua). http://doi.org/10.1109/ICIIBMS.2017.8279744

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้