



เว็บแอปพลิเคชันสำหรับตรวจประเมินระยะของโรคในผู้ป่วยโรคไตเรื้อรัง  
โดยใช้เทคโนโลยีปัญญาประดิษฐ์



วิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตาม  
หลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมชีวการแพทย์  
วิทยาลัยวิศวกรรมชีวการแพทย์

บัณฑิตวิทยาลัย มหาวิทยาลัยรังสิต  
ปีการศึกษา 2565



**WEB APPLICATION FOR SCREENING AND EVALUATION OF  
CHRONIC KIDNEY DISEASE USING ARTIFICIAL  
INTELLIGENCE TECHNOLOGY**

**BY**

**SAKUNA CHANGBOO**

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR  
THE DEGREE OF MASTER OF ENGINEERING  
IN BIOMEDICAL ENGINEERING  
COLLEGE OF BIOMEDICAL ENGINEERING**

**GRADUATE SCHOOL, RANGSIT UNIVERSITY**

**ACADEMIC YEAR 2022**

วิทยานิพนธ์ เรื่อง

เว็บแอปพลิเคชันสำหรับตรวจประเมินระยะของโรคในผู้ป่วยโรคไตเรื้อรัง  
โดยใช้เทคโนโลยีปัญญาประดิษฐ์

โดย

ศกุนา ช่างนู้

ได้รับการพิจารณาให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมชีวการแพทย์

มหาวิทยาลัยรังสิต

ปีการศึกษา 2565

รศ.ดร.ชูชาติ ปิณฑวิรุจน์  
ประธานกรรมการสอบ

รศ. นันทชัย ทองแป้น  
กรรมการ

ผศ.ดร.ณัฐพล ถนงค์ช่างแสง  
กรรมการและอาจารย์ที่ปรึกษาร่วม

ผศ.ดร.ศนิ นุญญกุล  
กรรมการและอาจารย์ที่ปรึกษา

บัณฑิตวิทยาลัยรับรองแล้ว

(ผศ.ร.ต.หญิง ดร.วรรณิ์ สุขสาตร)

คณบดีบัณฑิตวิทยาลัย

19 ตุลาคม 2565

Thesis entitled

**WEB APPLICATION FOR SCREENING AND EVALUATION OF  
CHRONIC KIDNEY DISEASE USING ARTIFICIAL  
INTELLIGENCE TECHNOLOGY**

by

SAKUNA CHANGBOO

was submitted in partial fulfillment of the requirements  
for the degree of Master of Engineering in Biomedical Engineering

Rangsit University  
Academic Year 2022

---

Assoc.Prof. Chuchart Pintavirooj, Ph.D.  
Examination Committee Chairperson

Assoc.Prof. Nuntachai Thongpance  
Member

---

Asst.Prof. Nuttapol Tanadchangsang, D.Eng.  
Member and Co-Advisor

Asst.Prof. Sani Boonyagul, Ph.D.  
Member and Advisor

Approved by Graduate School

(Asst.Prof.Plt.Off. Vanee Sooksatra, D.Eng.)

Dean of Graduate School

October 19, 2022

## กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงได้ด้วยความกรุณาจาก ผศ.ดร.ศนิ บุญญกุล อาจารย์ที่ปรึกษา วิทยานิพนธ์ และ อาจารย์ อนุชิต นีรภัย อาจารย์ที่ปรึกษาร่วมวิทยานิพนธ์ ที่ได้กรุณาให้คำปรึกษา แนะนำและแนะแนวทางอันเป็นประโยชน์ รวมทั้งการตรวจตราและแก้ไขเนื้อหาตลอดระยะเวลาที่ได้ทำการศึกษาค้นคว้าจนสำเร็จลุล่วงไปด้วยดี ขอกราบขอบพระคุณในความกรุณาเป็นอย่างสูง

ขอขอบพระคุณคณาจารย์ ในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมชีวการแพทย์ วิทยาลัยวิศวกรรมชีวการแพทย์ มหาวิทยาลัยรังสิต ทุกท่านที่ได้ให้ความรู้ในตลอดระยะเวลาการศึกษา ตลอดจนเจ้าหน้าที่ประจำสาขาที่คอยช่วยประสานงานเป็นอย่างดี

ขอขอบพระคุณบิดา มารดา และสมาชิกทุกคนในครอบครัว ที่กรุณาให้การอบรม สั่งสอน และเป็นกำลังใจ สนับสนุน และช่วยเหลือในทุกด้านจนกระทั่งประสบความสำเร็จได้ในทุกวันนี้

สกุณา ช่างบุ๋  
ผู้วิจัย



6305355 : สกุนา ช่างบู  
 ชื่อวิทยานิพนธ์ : เว็บแอปพลิเคชันสำหรับตรวจประเมินระยะของโรคในผู้ป่วย  
 โรคไตเรื้อรังโดยใช้เทคโนโลยีปัญญาประดิษฐ์  
 หลักสูตร : วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมชีวการแพทย์  
 อาจารย์ที่ปรึกษา : ผศ.ดร.ศนิ บุญญกุล  
 อาจารย์ที่ปรึกษาร่วม : ผศ.ดร.ณัฐพล ถนงค์ช่างแสง

### บทคัดย่อ

ด้วยอาการป่วยของคนไทยที่มีความชุกในลำดับต้นๆ ก็คือโรคไต ซึ่งเป็นโรคที่คุกคามสุขภาพคนไทยมาโดยตลอด และมีแนวโน้มจำนวนผู้ป่วยเพิ่มขึ้น การตรวจวินิจฉัยต้องใช้บุคลากรทางการแพทย์ที่มีอยู่จำกัด ทำให้ผู้ป่วยต้องรอและเสียโอกาสในการรักษา ด้วยความก้าวหน้าทางเทคโนโลยีทั้งการจัดเก็บ การประมวลผลข้อมูล และความสามารถของอุปกรณ์ที่มีประสิทธิภาพสูงขึ้น รวมทั้งเทคโนโลยีปัญญาประดิษฐ์ ทำให้การวิเคราะห์ข้อมูลและการทำนายโอกาส แนวโน้มในอนาคตมีความแม่นยำมากขึ้น งานวิจัยนี้วัตถุประสงค์เพื่อ 1) ศึกษาเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน 2) ศึกษาแบบจำลองที่ใช้ในการวิเคราะห์ระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน และ 3) พัฒนาเว็บแอปพลิเคชันสำหรับตรวจประเมินระยะของโรคในผู้ป่วยโรคไตเรื้อรัง โดยใช้เทคโนโลยีปัญญาประดิษฐ์ ตามหลักการเรียนรู้ของเครื่อง (Machine Learning: ML) ประเภท การเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยใช้ข้อมูลผู้ป่วยโรคเบาหวานจำนวน 1,528 คน จากคณะแพทยศาสตร์ วชิรพยาบาล ผลการวิจัยพบว่าอัลกอริทึมที่เหมาะสมคือ อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) และนำมาพัฒนาเว็บแอปพลิเคชันด้วยภาษาไพธอนแบบจำลองที่ได้มีค่าความถูกต้อง เท่ากับ 0.9760 ค่าความแม่นยำ เท่ากับ 0.9726 ค่าความระลึก เท่ากับ 0.9530 ค่าความถ่วงดุล เท่ากับ 0.9627 ตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ เท่ากับ 0.9701 เมื่อนำแบบจำลองและข้อมูลดังกล่าวมาสร้างเว็บแอปพลิเคชันพบว่า ผลการประเมินเว็บแอปพลิเคชัน มีค่าเฉลี่ยในภาพรวมเท่ากับ 4.45 อยู่ในระดับมาก

(วิทยานิพนธ์มีจำนวนทั้งสิ้น 89 หน้า)

คำสำคัญ: การเรียนรู้ของเครื่อง, การเรียนรู้แบบมีผู้สอน, โรคไตเรื้อรัง, โรคเบาหวาน

ลายมือชื่อนักศึกษา ..... ลายมือชื่ออาจารย์ที่ปรึกษา .....

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

6305355 : Sakuna Changboo  
 Thesis Title : Web Application for Screening and Evaluation of Chronic Kidney Disease Using Artificial Intelligence Teachnology  
 Program : Master of Engineering in Biomedical Engineering  
 Thesis Advisor : Asst.Prof. Sani Boonyagul, Ph.D.  
 Thesis Co-Advisor : Asst.Prof. Nuttapol Tanadchangsaeng, D.Eng.

**Abstract**

Among the most common illnesses in Thailand, a kidney disease is considered a continual threat to Thai people as seen from the increasing rate of patients with kidney diseases. Due to limitations on diagnosis and limited number of medical personnel, patients need to queue and sometimes lack opportunity to receive treatment. With the help of technological advancement in data collection, storage, and processing as well as the effectiveness of high-tech devices including artificial intelligence (AI), the diagnosis and prediction have become more valid. This research aimed to 1) conduct a comparative study on the effectiveness of the model for predicting kidney diseases in patient with diabetes, 2) investigate the model for predicting kidney diseases in patient with diabetes, and 3) develop a web application for evaluating stages of kidney diseases in patients with chronic kidney diseases (CKDs) with an application of AI according to the concept of supervised learning in Machine Learning (ML). The research investigated 1,528 patients with diabetes from Faculty of Medicine, Vajira Hospital. The result revealed that the best algorithm was Support Vector Machine (SVM) which was applied to the development of a python web application. The model had an accuracy of 0.9760, a precision of 0.9726, a recall of 0.9530, and an F-measure of 0.9627. The effectiveness of the model was 0.9701. The mean score of the web application evaluation was high ( $\bar{x}$  =4.45).

(Total 89 pages)

Keywords: Machine Learning (ML), Supervised Learning, Chronic kidney diseases (CKD), Diabetes

Student’s Signature ..... Thesis Advisor’s Signature .....

Thesis Co-Advisor’s Signature.....



## สารบัญ

	หน้า
กิตติกรรมประกาศ	ก
บทคัดย่อภาษาไทย	ข
บทคัดย่อภาษาอังกฤษ	ค
สารบัญ	ง
สารบัญตาราง	ฉ
สารบัญรูป	ช
<b>บทที่ 1</b>	
<b>บทนำ</b>	<b>1</b>
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์การวิจัย	3
1.3 สมมติฐานงานวิจัย	4
1.4 กรอบแนวคิดการวิจัย	4
1.5 ขอบเขตการวิจัย	5
1.6 นิยามศัพท์	6
<b>บทที่ 2</b>	
<b>ทบทวนวรรณกรรมที่เกี่ยวข้อง</b>	<b>8</b>
2.1 ทฤษฎีที่เกี่ยวข้อง	8
2.2 ผลงานวิจัยที่เกี่ยวข้อง	18
<b>บทที่ 3</b>	
<b>เครื่องมือและระเบียบวิธีการวิจัย</b>	<b>28</b>
3.1 เครื่องมือที่ใช้ในการวิจัย	28
3.2 ขั้นตอนและวิธีการวิจัย	28
3.3 ขั้นตอนและวิธีการทดสอบการวิจัย	34



## สารบัญ (ต่อ)

	หน้า
<b>บทที่ 4 ผลการวิจัย</b>	<b>48</b>
4.1 ผลการวิจัย	48
4.2 ผลการทดสอบ	56
<b>บทที่ 5 สรุปผลและข้อเสนอแนะ</b>	<b>75</b>
5.1 สรุปผลการทำวิจัย	75
5.2 การอภิปรายผลการจัดทำวิจัย	77
5.3 ข้อเสนอแนะ	77
<b>บรรณานุกรม</b>	<b>78</b>
<b>ภาคผนวก</b>	<b>79</b>
<b>ภาคผนวก ก</b> ตัวอย่างแบบประเมินความพึงพอใจ	81
<b>ภาคผนวก ข</b> การนำเสนอผลงานในงานประชุมวิชาการ	85
	The 17 UTCC National Graduate Research Conference
<b>ประวัติผู้วิจัย</b>	<b>89</b>

## สารบัญตาราง

ตารางที่	หน้า
2.1	9
2.2	21
2.3	23
2.4	25
3.1	34
3.2	40
3.3	47
4.1	50
4.2	57
4.3	64
4.4	69
4.5	73
4.6	74
4.7	74

## สารบัญรูป

รูปที่		หน้า
		5
1.1	กรอบแนวคิดการวิจัย	6
2.1	ภาพประกอบการเรียนรู้ของเครื่อง	11
2.2	แผนผังแสดงประเภทการเรียนรู้ของเครื่อง	13
2.3	แสดงลักษณะการทำงานของ Bangging	16
2.4	แสดงลักษณะการทำงานของ Boosting	17
2.5	แสดงวิธีการซัพพอร์ตเวกเตอร์แมชชีน	19
2.6	แสดงวิธีการค้นหาเพื่อนบ้านใกล้ที่สุด	20
3.1	แสดงขั้นตอนและวิธีการวิจัย	29
3.2	แสดงสถาปัตยกรรมของระบบวิเคราะห์และทำนายความเสี่ยงโรคไตเรื้อรัง	30
3.3	แสดงระบบการวิเคราะห์และทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน	32
3.4	แสดงการนำข้อมูลเข้าระบบ	35
3.5	แสดงข้อมูลที่ใช้ในการทดสอบระบบ	41
3.6	แสดงการเก็บแบบจำลอง	46
3.7	แสดงการสร้างส่วนต่อประสานกับผู้ใช้งาน	46
4.1	แสดงข้อมูลที่น่าเข้าสู่การวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน	49
4.2	แสดงกราฟเปรียบเทียบประสิทธิภาพด้านความถูกต้องของแต่ละอัลกอริทึม	51
4.3	แสดงกราฟเปรียบเทียบประสิทธิภาพด้านความแม่นยำของแต่ละอัลกอริทึม	51
4.4	แสดงกราฟเปรียบเทียบประสิทธิภาพด้านค่าเรียกกลับของแต่ละอัลกอริทึม	52
4.5	แสดงกราฟเปรียบเทียบประสิทธิภาพด้านค่าความถ่วงดุลของแต่ละอัลกอริทึม	53

สารบัญรูป (ต่อ)

รูปที่		หน้า
4.6	แสดงกราฟเปรียบเทียบประสิทธิภาพด้วยค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ ROC Curve	53
4.7	แสดงข้อมูลที่น่าเข้าสู่การวิเคราะห์ระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน	55
4.8	แสดงลักษณะและความมีอยู่ของข้อมูลนำเข้าระบบ ชุดที่ 1	55
4.9	แสดงความสัมพันธ์ของข้อมูลแต่ละตัว	56
4.10	แสดงค่า Confusion Matrix ของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน	58
4.11	กราฟแสดงผลการวัดประสิทธิภาพของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน	59
4.12	แสดงพื้นที่ใต้กราฟ ROC Curve ของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน โดยแบ่งผลลัพธ์เป็น 6 กลุ่ม	60
4.13	แสดงข้อมูลที่น่าเข้าสู่การทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน โดยใช้ข้อมูลของผู้ป่วยโรคเบาหวาน	61
4.14	แสดงลักษณะและความมีอยู่ของข้อมูลนำเข้าระบบ ชุดที่ 2	62
4.15	แสดงความสัมพันธ์ของข้อมูล	63
4.16	แสดงค่า Confusion Matrix ของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน	65
4.17	กราฟแสดงผลการวัดประสิทธิภาพของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน	66
4.18	แสดงพื้นที่ใต้กราฟ ROC Curve ของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน โดยแบ่งผลลัพธ์เป็น 6 กลุ่ม	67
4.19	แสดงหน้าจอบีเอ็มเอพพลิเคชันวิเคราะห์ระยะการเกิดโรคไตเรื้อรังและทำนายระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน	68
4.20	แสดงผลการวิเคราะห์โรคไตเรื้อรังสำหรับผู้ป่วยเบาหวาน	70
4.21	แสดงผลการทำนายโรคไตเรื้อรังสำหรับผู้ป่วยเบาหวาน	72

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ไตเป็นอวัยวะที่มีหน้าที่หลักในการกำจัดของเสียออกจากร่างกาย ปรับสมดุลน้ำ เกลือแร่ และกรดต่าง รวมทั้งสร้างฮอร์โมนเรนิน (Renin hormone) ช่วยในการควบคุมความดันโลหิตและการดูดซึมเกลือแร่ที่ไต ฮอร์โมนอีริโทรพอยอีติน (Erythropoietin Hormone) ช่วยให้ไขกระดูกสร้างเม็ดเลือดแดง วิตามินดีช่วยควบคุมการดูดซึมแคลเซียมจากอาหารและช่วยในการเสริมสร้างกระดูก ดังนั้นหากไตเสื่อมลงจนเสียหน้าที่จะส่งผลกระทบต่อระบบต่างๆ ของร่างกาย โดยธรรมชาติเมื่ออายุ 40 ปี ขึ้นไป ไตจะเริ่มเสื่อมหน้าที่ปีละ 1-2 % โดยโรคไตเรื้อรังจะแบ่งออกเป็น 5 ระยะ ซึ่งจะแบ่งตามระดับค่าอัตราการกรองของไต (estimated Glomerular Filtration Rate: eGFR) และปริมาณโปรตีนในปัสสาวะ (Albuminuria) ซึ่งทั้งสองปัจจัยนี้ มีผลสำคัญต่อการดำเนินโรคเข้าสู่ไตวายระยะสุดท้าย เมื่อแพทย์ตรวจพบว่าผู้ป่วยที่มี eGFR น้อยกว่า 60 มล./นาที/1.73 ตร.ม. แสดงว่าผู้ป่วยมีความเสี่ยงที่จะเป็นโรคไตเรื้อรัง โดยแพทย์จะตัดสินใจร่วมกับปัจจัยอื่นๆ ด้วย เช่น ปัจจัยที่มีผลต่อความไวในการเกิดโรค ปัจจัยที่มีผลต่อการเกิดโรคไตเรื้อรัง ปัจจัยที่มีผลต่อการดำเนินโรคของโรคไตเรื้อรัง และ ปัจจัยที่มีผลต่อการนำไปสู่โรคไตระยะสุดท้าย

สถานการณ์โรคไตในปัจจุบันกำลังเป็นปัญหาใหญ่ระดับโลก คนไทยมีแนวโน้มป่วยเพิ่มขึ้น สาเหตุส่วนใหญ่ร้อยละ 70 เกิดจากเบาหวานและความดันโลหิตสูง ซึ่งมีสถิติผู้ป่วยรวมเกือบ 15 ล้านคน ผลที่ตามมาคือมีภาวะไตเสื่อมและไตเสื่อมเร็วขึ้นหากปฏิบัติตัวไม่ถูกต้อง จากข้อมูลพบว่าคนไทยป่วยเป็นโรคไตเรื้อรังร้อยละ 17.6 ของประชากร หรือประมาณ 8 ล้านคน เป็นผู้ป่วยระยะสุดท้าย 2 แสนคน ป่วยเพิ่มปีละกว่า 7,800 ราย ส่วนการผ่าตัดเปลี่ยนไตทำได้เพียงปีละ 500 ราย จึงเน้นการชะลอความเสื่อมของไตเพื่อให้เข้าสู่ระยะที่ต้องล้างไตช้าลง ในปี 2563 สำนักงานหลักประกันสุขภาพแห่งชาติ ได้จัดสรรงบประมาณการบำบัดทดแทนโรคไตวายเรื้อรังจำนวนกว่า 9,405 ล้านบาท โดยคาดประมาณว่าจะมีผู้ป่วยในระบบหลักประกันสุขภาพแห่งชาติ 61,948 ราย แบ่งเป็นผู้ป่วยล้างไตผ่านช่องท้องจำนวน 31,047 ราย การฟอกเลือดจำนวน 28,546 ราย

และการผ่าตัดเปลี่ยนไตจำนวน 172 ราย ผู้ป่วยที่ต้องได้รับยากดภูมิคุ้มกัน 2,183 ราย ซึ่งล่าสุดข้อมูล ณ วันที่ 30 กันยายน 2562 มีผู้ป่วยทั้งสิ้น 59,830 คนแบ่งเป็นผู้ป่วยที่ได้รับการล้างไตผ่านช่องท้องแล้ว จำนวน 30,627 ราย การฟอกเลือดแล้ว 26,633 ราย ผ่าตัดเปลี่ยนไต 205 ราย และได้รับยากดภูมิคุ้มกันจำนวน 2,365 ราย ทั้งนี้ โดยเฉลี่ยมีค่าใช้จ่ายในการบำบัดทดแทนไตประมาณ 200,000 บาทต่อคนต่อปี นพ.ประพนธ์ ตั้งศรีเกียรติกุล 2020

ในช่วงที่ผ่านมาไม่นานนี้ปัญญาประดิษฐ์ (Artificial Intelligence: AI) ได้เข้ามามีบทบาทสำคัญกับการปฏิวัติทางสาธารณสุข เนื่องจากปริมาณข้อมูลที่ทางการแพทย์มีมากขึ้นมหาศาลในช่วงไม่กี่ปีที่ผ่านมา ความสามารถด้านฮาร์ดแวร์ (Hardware) ในการประมวลผลข้อมูลของคอมพิวเตอร์ที่เพิ่มมากขึ้น และการสนับสนุนทางการศึกษาเกี่ยวกับปัญญาประดิษฐ์ ของภาครัฐและเอกชน หนึ่งในอัลกอริทึมทางการเรียนรู้ของเครื่อง (Machine Learning) ที่ได้รับความสนใจและเป็นรากฐานสำคัญของเทคโนโลยีปัญญาประดิษฐ์ เกินกว่าครึ่งหนึ่งในปัจจุบันคือ การเรียนรู้เชิงลึก (Deep Learning) ซึ่งมีรากฐานที่ประกอบไปด้วยโครงข่ายประสาทเทียม (Neural Network) ที่ถูกคิดค้นตั้งแต่ปี ค.ศ. 1943 แต่เนื่องจากคอมพิวเตอร์ในอดีตยังมีข้อจำกัดด้านประสิทธิภาพ จึงยังไม่สามารถเรียนรู้ข้อมูลปริมาณมหาศาล หรือสร้างแบบจำลองได้ในระยะเวลาอันสั้น ประกอบกับข้อมูลดิจิทัลที่ยังมีจำนวนน้อยและไม่ได้บันทึกไว้อย่างเป็นระบบ จึงทำให้การนำข้อมูลไปใช้ประโยชน์ทำได้ยาก ส่งผลให้เทคโนโลยีโครงข่ายประสาทเทียมไม่ได้รับความนิยม จนกระทั่งในปี ค.ศ. 2012 ที่เทคโนโลยีการเรียนรู้เชิงลึก สามารถคว้าชัยชนะจากการแข่งขัน ImageNet Challenge ในการจำแนกประเภทของรูปภาพจำนวนกว่า 1,000,000 รูปออกเป็นประเภทต่าง ๆ ได้กว่า 1,000 ประเภทโดยมีความผิดพลาดต่ำที่สุดเป็นประวัติการณ์ ส่งผลให้การเรียนรู้เชิงลึก เป็นที่สนใจและได้รับการศึกษาค้นคว้ากันเป็นวงกว้าง ปัจจุบันเทคโนโลยีการเรียนรู้เชิงลึก ถูกนำมาใช้ประโยชน์หลากหลาย ตั้งแต่การวิเคราะห์ตลาดหุ้น การแยกแยะใบหน้าจากรูปภาพ ทำนายการจราจร รวมถึงด้านการแพทย์

ในด้านการแพทย์ปัญญาประดิษฐ์ มีประโยชน์ต่อผู้มีส่วนเกี่ยวข้อง 4 กลุ่ม คือ ผู้ป่วย ผู้ให้บริการทางการแพทย์ บุคคลทั่วไป และผู้พัฒนาฯ สำหรับผู้ป่วยจะได้ประโยชน์จากปัญญาประดิษฐ์ ในแง่ของการเข้าใจทำนายโรคที่ตนเองเป็น ทำให้สามารถเปลี่ยนแปลงการรักษาได้เหมาะสมกับตนเองมากขึ้น สำหรับแพทย์หรือเจ้าหน้าที่สถานพยาบาล ปัญญาประดิษฐ์ จะช่วยแบ่งเบาภาระงาน ลดต้นทุนการวินิจฉัยโรค และช่วยอำนวยความสะดวกในสถานพยาบาลให้เป็นไปอย่างรวดเร็ว เช่น ปัญญาประดิษฐ์ในโรงพยาบาลที่คอยจัดคิวเพื่อให้ผู้ป่วยได้พบแพทย์ห้องฉุกเฉิน



ตามลำดับความเร่งด่วน หรือ ปัญญาประดิษฐ์ที่ช่วยวินิจฉัยโรคมะเร็งผิวหนังจากภาพถ่าย ซึ่งมีหลักฐานแสดงให้เห็นว่ามีความแม่นยำในการวินิจฉัยโรคเทียบเคียงกับแพทย์ผู้เชี่ยวชาญด้านโรคผิวหนังตัวอย่างงานวิจัยทางด้าน ปัญญาประดิษฐ์ ที่มีชื่อเสียงและได้รับการยอมรับในแวดวงนักวิจัยและวงการแพทย์ ในด้านการพัฒนาและคิดค้นยาใหม่ ปัญญาประดิษฐ์เข้ามามีบทบาทตั้งแต่ขั้นตอนการคิดค้นยา โดยช่วยวิเคราะห์ส่วนประกอบของโมเลกุลและสมบัติของสารประกอบ จนถึงการทำนายประสิทธิภาพและผลข้างเคียงของยาชนิดใหม่ ด้วยปัจจัยทั้งหมดที่กล่าวมา อนาคตทางการแพทย์มีการเปลี่ยนแปลงจากยุคของเวชปฏิบัติอิงหลักฐาน (Evidence-Based Medicine: EBM) ไปเป็นเวชศาสตร์ที่การพัฒนาถูกขับเคลื่อนโดยข้อมูลมหัต (Data-Driven Medicine) ซึ่งปัจจุบันนี้โรงพยาบาลต่าง ๆ ทั่วโลก รวมถึงในประเทศไทยเริ่มให้ความสำคัญกับการบันทึกข้อมูลทุกอย่างในรูปแบบดิจิทัลทำให้แนวโน้มในอนาคตปัญญาประดิษฐ์จะเข้ามามีบทบาทในทางการแพทย์มากยิ่งขึ้น ไม่ว่าจะเป็นด้านการวินิจฉัย การรักษา หรือการทำนายโรค

ดังนั้นจากปัญหาและความก้าวหน้าทางเทคโนโลยีที่กล่าวมาข้างต้น จึงทำให้เกิดแนวคิดในการพัฒนาระบบวิเคราะห์และทำนายความเสี่ยงการเกิดโรคไตเรื้อรัง โดยใช้ปัญญาประดิษฐ์ ที่จะสามารถวิเคราะห์และทำนายโอกาสที่จะเกิดโรคไตเรื้อรัง และในรายที่เป็นโรคไตเรื้อรังอยู่แล้วสามารถทำนายระยะเวลาการเกิดไตวายระยะสุดท้าย นอกจากนี้ยังช่วยวิเคราะห์และทำนายผลว่าปัจจัยใดที่ทำให้ผู้ป่วยมีค่าอัตราการกรองของไตลดลง เพื่อให้แพทย์ทราบว่าผู้ป่วยจะมีความเสี่ยงที่จะเป็นโรคไตเรื้อรังในระยะใด ทำให้แพทย์สามารถวางแผนการรักษาได้อย่างมีประสิทธิภาพ โดยการพัฒนาปัญญาประดิษฐ์ ในรูปแบบแอปพลิเคชัน โดยขอบเขตของการพัฒนาระบบมุ่งเน้นไปที่ผู้ป่วยในกลุ่มที่เป็นเบาหวานเพื่อหาความสัมพันธ์ของการเกิดโรคไตเรื้อรังในผู้ป่วยเบาหวาน รวมถึงระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน

## 1.2 วัตถุประสงค์งานวิจัย

1.2.1 ศึกษาเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน

1.2.2 ศึกษาแบบจำลองที่ใช้ในการวิเคราะห์ระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน



1.2.3 พัฒนาเว็บแอปพลิเคชันสำหรับตรวจประเมินระยะของโรคในผู้ป่วยโรคไตเรื้อรัง โดยใช้เทคโนโลยีปัญญาประดิษฐ์

### 1.3 สมมติฐานของการวิจัย

สามารถนำข้อมูลของผู้ป่วยโรคไตเรื้อรังและข้อมูลผู้ป่วยเบาหวานเพื่อวิเคราะห์และทำนายการเกิดโรคไตเรื้อรังในผู้ป่วยเบาหวาน และการเกิดภาวะไตวาย

#### 1.3.1 ตัวแปรต้น

1.3.1.1 ข้อมูลผู้ป่วยโรคเบาหวาน ได้แก่

- |             |   |
|-------------|---|
| 1) ID       | คือ รหัสผู้ป่วย   |
| 2) eGFR     | คือ อัตราการกรองของไต (ml/min/1.73 m <sup>2</sup> )         |
| 3) Age      | คือ อายุ (ปี)   |
| 4) Duration | คือ ระยะเวลาที่ผู้ป่วยเป็นโรครมาแล้วกี่ปี (ปี)              |
| 5) BMI      | คือ ดัชนีมวลกาย (kg/ m <sup>2</sup> )                       |
| 6) SBP      | คือ ความดันสูงสุดของหัวใจห้องล่างบีบตัว (mmHg)              |
| 7) DBP      | คือ ความดันเลือดที่ต่ำที่สุดขนาดหัวใจห้องล่างคลายตัว (mmHg) |
| 8) FBS      | คือ การตรวจระดับน้ำตาลในเลือดหลังจากอดอาหาร 8 ชั่วโมง       |
| (mg/dL)     |   |
| 9) HbA1c    | คือ การตรวจระดับน้ำตาลในเลือดตลอดระยะเวลา 2 – 3 เดือนที่    |
| ผ่านมา (%)  |   |
| 10) CHO     | คือ คอเลสเตอรอล (dL)  |
| 11) TG      | คือ ไขมันในเลือด (mg/dL)                                    |
| 12) HDL     | คือ ไขมันชนิดดี (mg/dL)                                     |
| 13) LDL     | คือ ไขมันชนิดไม่ดี (mg/dL)                                  |

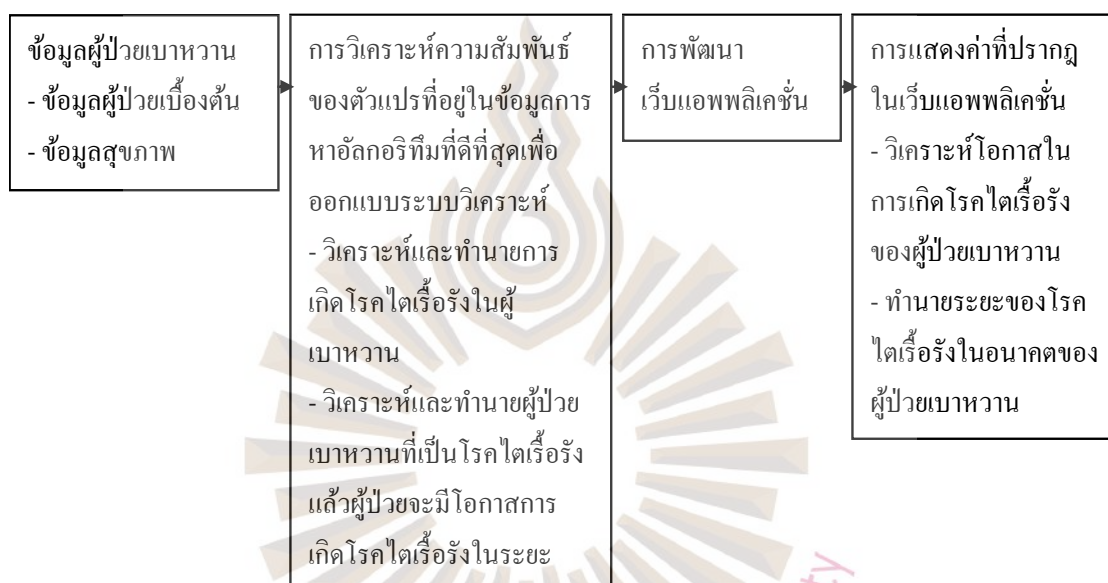
#### 1.3.2 ตัวแปรตาม

1.3.2.1 ระบบวิเคราะห์

### 1.3.3 ตัวแปรควบคุม

1.3.3.1 กลุ่มตัวอย่างซึ่งเป็นข้อมูลจากผู้ป่วยเบาหวาน จำนวน 1,528 ราย จากคณะแพทยศาสตร์วชิรพยาบาล มหาวิทยาลัยนวมินทราชิราช

## 1.4 กรอบแนวคิดการวิจัย



รูปที่ 1.1 กรอบแนวคิดการวิจัย

## 1.5 ขอบเขตการวิจัย

1.5.1 หาความสัมพันธ์ของตัวแปรจากข้อมูลสุขภาพของผู้ป่วยเบาหวาน

1.5.2 จากความสัมพันธ์ของตัวแปรนำมาวิเคราะห์เพื่อหาอัลกอริทึมของกระบวนการปัญญาประดิษฐ์สำหรับทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน

1.5.3 พัฒนาเว็บแอปพลิเคชันสำหรับตรวจประเมินระยะของโรคในผู้ป่วยโรคไตเรื้อรัง โดยใช้เทคโนโลยีปัญญาประดิษฐ์

1.5.4 ทดสอบระบบเว็บแอปพลิเคชันที่พัฒนาขึ้น โดยผู้ใช้งานจริงจำนวน 6 ราย

## 1.6 นิยามศัพท์

**โรคไตเรื้อรัง (Chronic Kidney Disease: CKD)** คือ ภาวะที่ไตมีความผิดปกติติดต่อกันเกิน 3 เดือน ซึ่งผู้ป่วยอาจจะมีอัตราการกรองของไต (Estimated Glomerular Filtration Rate: eGFR) ผิดปกติหรือไม่ก็ได้โดยสามารถตรวจความผิดปกติทางพยาธิสภาพหรือตัวบ่งชี้ว่าไตถูกทำลายจากความผิดปกติของเลือดหรือปัสสาวะ การตรวจทางรังสีหรืออัตราการกรองของไต (eGFR) ลดลงน้อยกว่า 60 มิลลิลิตร/นาที/พื้นผิวร่างกาย 1.73 ตารางเมตร เป็นเวลามากกว่าหรือเท่ากับ 3 เดือน

**โรคเบาหวาน** คือ โรคที่เซลล์ร่างกายมีความผิดปกติในขบวนการเปลี่ยนน้ำตาลในเลือดให้เป็นพลังงาน โดยขบวนการนี้เกี่ยวข้องกับอินซูลิน (Insulin) ซึ่งเป็นฮอร์โมนที่สร้างจากตับอ่อนเพื่อใช้ควบคุมระดับน้ำตาลในเลือด เมื่อน้ำตาลไม่ได้ถูกใช้จึงทำให้ระดับน้ำตาลในเลือดสูงขึ้นกว่าระดับปกติ

**Graphics Processing Unit (GPU)** คือ อุปกรณ์ที่มีได้ทั้งที่เป็นการ์ดหรือเป็นส่วนหนึ่งของแผงเมนบอร์ดก็ได้แต่ในปัจจุบันการ์ดแสดงผลส่วนใหญ่อยู่ในรูปของการ์ด หน้าที่หลักของ GPU คือช่วยในการประมวลการทำงานในด้านภาพกราฟฟิกบนหน้าจอคอมพิวเตอร์ให้มีประสิทธิภาพมากยิ่งขึ้น โดยมีหลักการทำงานคล้ายกับ CPU แต่จะแตกต่างกันตรงที่การ์ดแสดงผลยุคเก่าจะทำหน้าที่แปลงข้อมูลดิจิทัลเป็นสัญญาณเท่านั้น

**ปัญญาประดิษฐ์** คือ ความฉลาดเทียมที่สร้างขึ้นให้กับสิ่งที่ไม่มีชีวิต ซึ่งเป็นการเรียนรู้เกี่ยวกับกระบวนการการคิด การกระทำ การให้เหตุผล การปรับตัวหรือการอนุมาน และการทำงานของสมอง แม้ว่าเดิมนั้นเป็นสาขาหลักในวิทยาการคอมพิวเตอร์ แต่แนวคิดหลาย อย่างในศาสตร์นี้ได้มาจากการปรับปรุงเพิ่มเติมจากศาสตร์อื่น

**การเรียนรู้ของเครื่อง (Machine Learning: ML)** คือ ถูกใช้งานเสมือนเป็นสมองของปัญญาประดิษฐ์ ซึ่งใช้การเรียนรู้ของเครื่องในการสร้างความฉลาด มนุษย์มีหน้าที่เขียนโปรแกรมให้ AI เรียนรู้จากข้อมูลเท่านั้น โดยการเรียนรู้ของเครื่องจะเรียนรู้จากสิ่งที่ส่งเข้าไปกระตุ้น แล้วจดจำเอาไว้เป็นมันสมองก่อนส่งผลลัพธ์ออกมาเป็นตัวเลข

**ระบบผู้เชี่ยวชาญ (Expert Systems: ES)** คือ การพัฒนาให้ระบบคอมพิวเตอร์มีความรู้ รู้จักใช้เหตุผลในการ วิเคราะห์ปัญหา โดยใช้ความรู้ที่มี หรือจากประสบการณ์ในการแก้ปัญหาหนึ่งไปแก้ไขปัญหาอื่นอย่างมีเหตุผล ระบบนี้จำเป็นต้องอาศัยฐานข้อมูล ซึ่งมนุษย์ผู้มีความรู้ความสามารถเป็นผู้กำหนดองค์ความรู้ ไว้ในฐานข้อมูลดังกล่าวเพื่อให้ระบบคอมพิวเตอร์สามารถ วิเคราะห์ปัญหาต่าง ๆ ได้จากฐานความรู้นั้น



## บทที่ 2

### ทบทวนวรรณกรรมที่เกี่ยวข้อง

#### 2.1 ทฤษฎีที่เกี่ยวข้อง

##### 2.1.1 โรคไตเรื้อรัง

สภาวะที่ไตถูกทำลายมีผลทำให้ความสามารถของไตในการทำงานลดลง เช่น การรักษาสมดุลของเหลวในร่างกาย การควบคุมน้ำหนักร่างกายและแร่ธาตุต่าง ๆ ในเลือด การกำจัดยาและพิษออกจากร่างกายและการหลั่งฮอร์โมนเข้าสู่กระแสเลือด เป็นต้น โดยสาเหตุที่ก่อให้เกิดโรคไตเรื้อรังได้แก่ เบาหวาน ความดันโลหิตสูงและโรคอ้วน รวมถึงสภาวะอื่น ๆ เช่น ไตอักเสบ โรคถุงน้ำในไต เป็นต้น

การทำงานของไตอาจมีประสิทธิผลลดลงทีละน้อย ทำให้ผู้ป่วยบางรายไม่สามารถทราบว่าเป็นโรคไตเรื้อรัง จนกระทั่งมีอาการรุนแรง ดังนั้นการตรวจพบและได้รับการรักษาตั้งแต่ระยะเริ่มแรก อาจทำให้โรคไตเรื้อรังคงตัวและสามารถหายได้แต่หากปล่อยไว้จนมีอาการมากอาจนำไปสู่ภาวะไตวาย ซึ่งเป็นภาวะที่ไตไม่สามารถทำงานได้ตามปกติอีกต่อไปทำให้ผู้ป่วยจำเป็นต้องได้รับการรักษาเช่น ฟอกเลือด หรือ เปลี่ยนไต

2.1.1.1 ปัจจัยเสี่ยงของโรคไตเรื้อรัง แบ่งออกเป็น 6 ชนิด คือ

- 1) โรคเบาหวาน
- 2) ความดันโลหิตสูง
- 3) ประวัติคนในครอบครัวเป็นโรคไตเรื้อรัง
- 4) อายุมาก
- 5) น้ำหนักเกิน
- 6) สูบบุหรี่

2.1.1.2 ระยะของโรคไตเรื้อรัง

โรคไตเรื้อรังจะแบ่งออกเป็น 5 ระยะ ซึ่งแบ่งตามระดับค่าอัตรากรองของไต

ตารางที่ 2.1 ระยะของโรคไตเรื้อรัง

ระยะของโรคไตเรื้อรัง	ค่าการทำงานของไต
ระยะที่ 1	ไตเริ่มเสื่อม ค่า eGFR 90 หรือมากกว่า
ระยะที่ 2	ไตเสื่อม ค่า eGFR 60 - 89
ระยะที่ 3a	ค่า eGFR ลดลงเล็กน้อย 45 - 59
ระยะที่ 3b	ค่า eGFR ลดลงปานกลาง 30 - 44
ระยะที่ 4	ค่า eGFR ลดลงมาก 15 - 29
ระยะที่ 5	ไตวาย ค่า eGFR น้อยกว่า 15

ที่มา: ปิยะธิดา จิงสมาน, ม.ป.ป.

### 2.1.2 โรคเบาหวาน (Diabetes Mellitus: DM)

เป็นภาวะที่ร่างกายมีน้ำตาลในเลือดสูงกว่าปกติเนื่องจากการขาดฮอร์โมนอินซูลิน (Insulin) หรือการดื้อต่อฮอร์โมนอินซูลิน ส่งผลให้กระบวนการดูดซึมน้ำตาลในเลือดให้เป็นพลังงานของเซลล์ในร่างกายมีความผิดปกติหรือทำงานได้ไม่เต็มประสิทธิภาพ จนเกิดน้ำตาลสะสมในเลือดปริมาณมาก หากปล่อยให้ร่างกายอยู่ในสภาวะนี้เป็นเวลานานจะทำให้อวัยวะต่าง ๆ เสื่อมเกิดโรคและอาการแทรกซ้อน โรคเบาหวานในระยะแรกจะไม่แสดงอาการผิดปกติ บางรายอาจตรวจพบโรคเบาหวานเมื่อพบภาวะแทรกซ้อนขึ้นแล้ว อาการของโรคเบาหวานแต่ละชนิดอาจมีความคล้ายกัน ซึ่งอาการที่พบส่วนใหญ่ คือ กระหายน้ำมาก ปากแห้ง ปัสสาวะบ่อย หิวบ่อย น้ำหนักลดหรือเพิ่มผิดปกติ สายตาพร่ามัว เห็นภาพไม่ชัด รู้สึกเหนื่อยง่าย มีอาการชาโดยเฉพาะมือและขาขาดแผลหายยาก เป็นต้น

### 2.1.3 ปัญญาประดิษฐ์

เป็นวิธีการที่ทำให้คอมพิวเตอร์ มีความสามารถคล้ายมนุษย์หรือ เลียนแบบพฤติกรรมของมนุษย์ ในการทำความเข้าใจ การเรียนรู้องค์ความรู้ต่างๆ เช่น การรับรู้และการเรียนรู้ การให้เหตุผล การแก้ปัญหา เช่น การเรียนรู้ของมนุษย์นั้น คือ จำ แล้ว คิด ตาม เช่น เด็กที่เห็นหน้าพ่อแม่ซ้ำ ๆ ทุกวัน และป้อนเสียงเรียก พ่อ แม่ ไปให้เด็ก เมื่อเวลาผ่านไปเด็กคนนั้นก็จะสามารถมองเห็นแล้วเรียกพ่อแม่ ออกมาได้โดยอัตโนมัติ สิ่งเร้าที่ใช้ฝึก (Train) ปัญญาประดิษฐ์ก็คือข้อมูลซึ่งในกรณีสอนเด็กให้เรียก พ่อแม่ ต้องใช้เวลานานในการฝึก เช่นเดียวกับ ปัญญาประดิษฐ์ที่ทำงาน โดยรวบรวมข้อมูล



ปริมาณมากมายมหาศาล ประมวลผลซ้ำๆ โดยการผ่านขั้นตอนการประมวลผล ด้วยซอฟต์แวร์ที่สามารถเรียนรู้จากรูปแบบและลักษณะของข้อมูลได้อย่างอัตโนมัติบนพื้นฐานทางทฤษฎี วิธีการและเทคโนโลยี โดยกลไกการทำงาน คือ ระบบประมวลผลทางคอมพิวเตอร์ และมีการเรียนรู้ของเครื่องเป็นส่วนประกอบ ซึ่งการเรียนรู้ของเครื่องนี้มีอัลกอริทึม ที่หลากหลาย ขึ้นอยู่กับปัญหา และข้อมูลที่ใช้ฝึกสอน เช่น การเรียนรู้เชิงลึกเป็นอัลกอริทึมที่เหมาะสมกับข้อมูลซับซ้อนขนาดใหญ่ วิธีการสุ่มป่าไม้ (Random Forest) เป็นอัลกอริทึมที่ใช้สำหรับปัญหาการเรียนรู้แบบมีผู้สอน (Supervised Learning) เป็นต้น โดยยกตัวอย่างให้เห็นภาพ เช่น ปัญญาประดิษฐ์จะเก็บข้อมูลของผู้ป่วยเนื้องอกในสมองและผลของการรักษาด้วยวิธีต่าง ๆ ไว้ทั้งหมด ทำให้ผู้ป่วยสามารถทราบได้ว่า การรักษาวิธีใดจะได้ผลกับเนื้องอกชนิดของตนเองมากที่สุด และในกรณีที่เลวร้ายที่สุดตนเองจะสามารถมีชีวิตอยู่ได้อีกนานเท่าไร ปัญญาประดิษฐ์ที่ช่วยวินิจฉัยโรคจอประสาทตาจากภาพถ่าย ซึ่งกำลังพัฒนาโดยบริษัทกูเกิล (Google) ปัญญาประดิษฐ์วินิจฉัยโรคมะเร็งผิวหนังจากกล้องส่องผิวหนัง (Dermoscope) ที่มีความแม่นยำเทียบเท่ากับอายุรแพทย์โรคผิวหนัง ปัญญาประดิษฐ์ที่สามารถคัดกรองวินิจฉัยโรคปอดจากภาพถ่ายเอกซเรย์รังสีทรวงอก โดยจะระบุตำแหน่งที่น่าสงสัยก่อนส่งให้รังสีแพทย์อ่านต่อไป หรือระบบอื่น ๆ ที่ใช้ ปัญญาประดิษฐ์วิเคราะห์ข้อมูลที่ไม่ใช่ภาพถ่ายเช่น ระบบที่สามารถทำนายได้ว่าผู้ป่วยจะมาตรวจที่โรงพยาบาลครั้งต่อไปด้วยอาการอะไร โดยวิเคราะห์จากข้อมูลการมาโรงพยาบาลครั้งก่อน ๆ ในเวชระเบียน และ ปัญญาประดิษฐ์ที่สามารถทำนายผู้ป่วยที่มีภาวะลมชักจากคลื่นไฟฟ้าสมอง เป็นต้น ส่วนประโยชน์ต่อบุคคลทั่วไป ปัญญาประดิษฐ์จะช่วยให้ความรู้และลดความเหลื่อมล้ำด้านการรักษาพยาบาลสำหรับประชาชน เช่น ปัญญาประดิษฐ์ ที่ให้ความรู้เกี่ยวกับโรคต่างๆ ซึ่งจะให้คำแนะนำเบื้องต้นเกี่ยวกับการรักษาอาการเจ็บป่วยและการปฏิบัติตัวเพื่อป้องกันโรค รวมถึงการฟื้นฟูภายหลังออกจากโรงพยาบาล

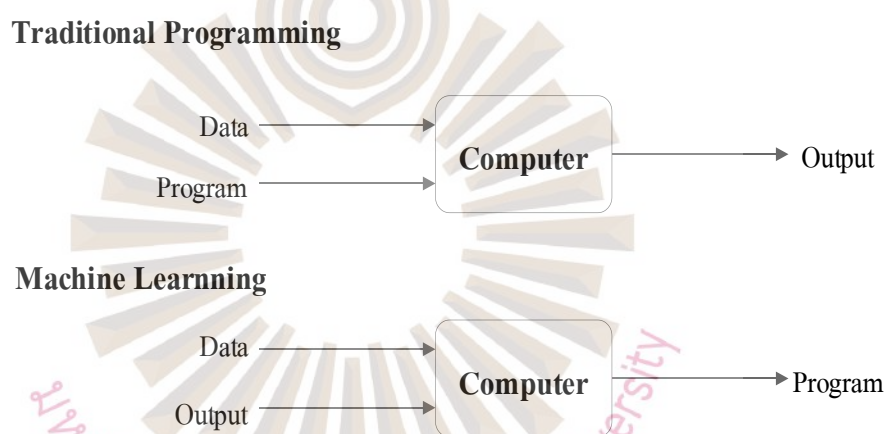
2.1.3.1 ปัญญาประดิษฐ์แบ่งตามความสามารถ 3 ระดับ คือ

- 1) ปัญญาประดิษฐ์เชิงแคบ (Narrow AI) มีความสามารถเฉพาะทางที่ดีกว่ามนุษย์ ซึ่งปัญญาประดิษฐ์ ณ ปัจจุบัน ยังอยู่ในระดับนี้
- 2) ปัญญาประดิษฐ์ทั่วไป (General AI) มีความสามารถระดับเดียวกับมนุษย์ ทำทุกอย่างที่มนุษย์ทำได้และมีประสิทธิภาพใกล้เคียงกัน
- 3) ปัญญาประดิษฐ์แบบเข้ม (Strong AI) มีความสามารถเหนือมนุษย์หลากหลายด้าน



### 2.1.4 การเรียนรู้ของเครื่อง

ส่วนการเรียนรู้ของเครื่อง ถูกใช้งานเสมือนเป็นสมองของปัญญาประดิษฐ์ ซึ่งปัญญาประดิษฐ์ใช้ การเรียนรู้ของเครื่องในการสร้างความฉลาด ซึ่งแตกต่างจากการเขียนโปรแกรมโดยทั่วไป เพราะผู้พัฒนาระบบ (Programming) จะใส่ ข้อมูลและชุดคำสั่ง (Program) เข้าไปเพื่อให้ได้ผลลัพธ์หรือเอาต์พุต (Output) การเรียนรู้ของเครื่องไม่ได้มีชุดคำสั่ง แต่เป็นการใส่ข้อมูล (Data) และผลลัพธ์เข้าไป เพื่อให้หาชุดคำสั่ง ที่จะนำไปเป็นผลลัพธ์ในอนาคตได้ว่าใส่ข้อมูลลักษณะแบบนี้ผลลัพธ์จะเป็นอะไร โดยการใช้ข้อมูลและการประยุกต์ใช้สถิติขั้นสูงเพื่อเรียนรู้ให้สามารถระบุรูปแบบข้อมูลและทำนายจากรูปแบบเหล่านั้นได้สามารถอธิบายลักษณะการเรียนรู้ของเครื่องได้ ดังรูปที่ 2.1



รูปที่ 2.1 ภาพประกอบการเรียนรู้ของเครื่อง

ที่มา: Kumar, 2019

#### 2.1.4.1 การเรียนรู้ของเครื่อง แบ่งออกเป็น 3 ประเภท คือ

##### 1) การเรียนรู้แบบมีผู้สอน (Supervised Learning)

การเรียนรู้ของเครื่องประเภทการเรียนรู้แบบมีผู้สอน และขั้นตอนการเรียนรู้ของเครื่องนั้นจะเลือกใช้อัลกอริทึมที่เหมาะสมกับลักษณะของปัญหาที่ต้องการแก้ไข โดยให้เรียนรู้จากข้อมูลนำเข้า (Input Data) ซึ่งประกอบไปด้วยข้อมูลคุณลักษณะ (Feature) และข้อมูลผลลัพธ์ (Label) จากนั้นจึงทำการทำนายผลลัพธ์ (Output Data) โดยข้อมูลที่ใช้สำหรับการเรียนรู้นั้นจะต้องมีคุณสมบัติทั้งสนับสนุนการเรียนรู้และการแก้ปัญหาเรียกว่าเวกเตอร์ของคุณลักษณะ (Feature

Vector) เพื่อสร้างแบบจำลอง (Model) สำหรับการเรียนรู้แบบมีผู้สอน จะมีการแบ่งชุดข้อมูล ออกเป็นชุดฝึก (Training Set) และ ชุดทดสอบ (Testing Set) เสมอ การแบ่งกลุ่มเป็น ชุดฝึกกับ ชุดทดสอบ ในการ เรียนรู้เพื่อป้องกันการเกิด กรณีการที่แบบจำลองมีความแม่นยำสูงมากในตอน ทดลอง แต่ พอเอาไปใช้งานจริงกลับได้ผลลัพธ์ที่ไม่ดี (Overfit) สาเหตุอาจเกิดจากการใช้ต้นแบบใน การเรียนรู้ และ วัตถุประสงค์การเรียนรู้เฉพาะกลุ่มชุดข้อมูลที่ใช้เป็นต้นแบบ และ ไม่มีการทดสอบกับชุด ข้อมูลอื่น การประเมินผลลัพธ์ของการเรียนรู้แบบมีผู้สอน มักจะใช้ความแม่นยำ (Accuracy) เป็น หลัก โดยแต่ละ โจทย์อาจ มีวิธีการวัดความแม่นยำที่ไม่เหมือนกัน เช่น การคิดเปอร์เซ็นต์ค่าความ คลาดเคลื่อนเฉลี่ย (Mean Absolute Percentage Error: MAPE) หรือ ค่าความคลาดเคลื่อนยกกำลัง สอง (Mean Squared Error: MSE) ทำให้มุมมองในการใช้งานจะแตกต่างกัน

### 1.1) การเรียนรู้แบบมีผู้สอนแบ่งออกเป็น 2 ประเภท คือ

(1) การจำแนกข้อมูล (Classification) คือ เป้าหมาย (Target) ที่เป็น ประเภท เช่น Yes หรือ No, 0 หรือ 1

(2) การถดถอย (Regression) คือ เป้าหมายที่มีค่าอย่างต่อเนื่องไม่ สามารถแบ่งประเภทได้ เช่น ราคาของสินค้า

### 2) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

การเรียนรู้ของเครื่องประเภทการเรียนรู้แบบไม่มีผู้สอน เป็นการใช้อัลกอริทึมแบบ ไม่มีการสอนและ ไม่มีการระบุผล (Target Variable) มีหลักการเรียนรู้จากข้อมูลนำเข้า ซึ่งประกอบ ไปด้วยข้อมูลคุณลักษณะ เท่านั้นและหาความสัมพันธ์จากข้อมูลด้วยตัวเอง

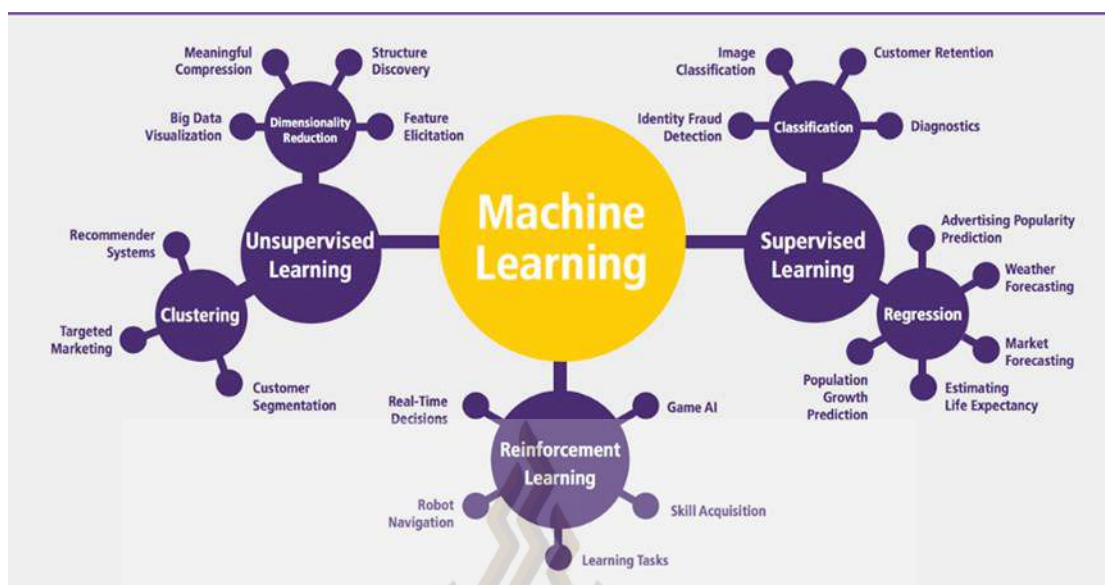
### 2.1) การเรียนรู้แบบไม่มีผู้สอนแบ่งออกเป็น 2 ประเภท คือ

(1) การจัดกลุ่มของข้อมูล (Clustering)

(2) การลดจำนวนมิติเพื่อบีบอัดข้อมูล (Dimensionality Reduction)

### 3) การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning)

การเรียนรู้แบบเสริมกำลัง เป็นการใช้อัลกอริทึมแบบหนึ่ง มีหลักการทำงาน เหมือนกับการที่มนุษย์เรียนรู้โดยใช้ประสบการณ์ การลองผิด-ลองถูก และ การเรียนรู้ผลดี ผลเสีย ของวิธีการแก้ปัญหาหนึ่ง ๆ เพื่อหาวิธีการแก้ปัญหาที่ได้ผลลัพธ์ที่ดีที่สุดกับปัญหานั้น โดยที่ หลักการคือการเรียนรู้ของ ผู้กระทำ (Agent) ที่เกิดจากปฏิสัมพันธ์แบบการลองผิด ลองถูก และ การ เรียนรู้ผลดี ผลเสีย ระหว่างผู้กระทำ กับสิ่งแวดล้อม โดยที่ผู้กระทำจะรับรู้ถึงสถานการณ์ของ สิ่งแวดล้อมผ่านสถานะ (State) และเลือกการกระทำ (Action) ที่ส่งผลต่อ สิ่งแวดล้อมโดยหวังว่าจะ ได้ตัวประเมินผลลัพธ์ (Reward) ที่ดีที่สุด รวมทั้งเรียนรู้จากข้อผิดพลาดในอดีตที่เกิดขึ้น สามารถ แสดงแผนผังประเภทการเรียนรู้ของเครื่องได้ดังรูปที่ 2.2



รูปที่ 2.2 แผนผังแสดงประเภทการเรียนรู้ของเครื่อง  
ที่มา: Samad, 2019

### 2.1.5 การวิเคราะห์ข้อมูล (Data Analysis)

เป็นทักษะที่กำลังเป็นที่ต้องการมากตั้งแต่การเกิดขึ้นของเทคโนโลยี เช่น อินเทอร์เน็ต และ โทรศัพท์มือถือ รวมไปถึงแท็บเล็ต จึงทำให้มีความต้องการในงานสายนี้มากขึ้นเรื่อย ๆ การเติบโตของแวดวงธุรกิจนี้ ผลักดันให้องค์กรตระหนักถึงข้อมูลจำนวนมากที่ถูกส่งไปมาทุก ๆ วินาที องค์กรสามารถติดตามและรวบรวมข้อมูลต่าง ๆ ที่จะเป็นประโยชน์ต่อตัวเองหรือทำให้ได้รับผลตอบแทนทางธุรกิจมากขึ้น โดยมีการทำนายว่าวิทยาการข้อมูลจะมีอิทธิพลต่อเศรษฐกิจโลกในอนาคตอย่างแน่นอน วิทยาการข้อมูลประกอบไปด้วย 2 ส่วน ดังนี้

2.1.5.1 การวิเคราะห์สถานการณ์การตลาด (Market Analysis) ความเข้าใจเกี่ยวกับการวิเคราะห์การตลาด จึงเป็นสิ่งที่ช่วยกำหนดเป้าหมายทางการตลาดและแบบแผนการตลาดได้ หากได้ทำความเข้าใจกับสิ่งเหล่านี้ การวางแผนทางการตลาดจะง่ายขึ้นและจะไม่เสียเงิน โดยใช่เหตุ กับการตลาดในโลกยุคใหม่ที่เติบโตจนล้นโลกแล้วก็ได้ ผู้ประกอบการควรใช้ความรอบคอบในการวางแผนสู่การตอบสนองผู้บริโภค

2.1.5.2 การทำเหมืองข้อมูล (Data Mining) หมายถึงการใช้ความรู้ทางด้านของคณิตศาสตร์ สมการ และ สถิติ ในการนำแยกข้อมูลทันสมัย และรูปแบบข้อมูลออกมาเพื่อใช้ในการโฆษณา และการตลาดอย่างมีประสิทธิภาพ รวมถึงการเริ่มทำธุรกิจออนไลน์ ห่วงโซ่อุปทานและ

ส่วนประกอบทางธุรกิจอื่น ๆ องค์กรสามารถปรับเปลี่ยนกระบวนการเพื่อเพิ่มการผลิต และ ประสิทธิภาพได้ จากการนำข้อมูลมาใช้ การทำเหมืองข้อมูลจึงเป็นสิ่งที่ช่วยเพิ่มผลประโยชน์ และทำให้ธุรกิจไปได้ราบรื่นยิ่งขึ้น การทำเหมืองข้อมูลแบ่งออกเป็น 3 ขั้นตอนย่อย ดังนี้

1) ขั้นก่อนดำเนินงาน (Preprocessing) ก่อนจะเริ่มการดำเนินงานทำเหมือง ข้อมูลได้นั้น จะต้องมีการจัดวางชุดข้อมูลก่อนเพราะเหมืองข้อมูลจะสามารถหาได้เพียงรูปแบบที่ กำหนดในข้อมูลก่อนแล้ว ชุดข้อมูลที่กำหนดจะต้องมีขนาดใหญ่พอที่จะมองเห็นรูปแบบ และ สัน กระจับพอที่จะชุดข้อมูลภายในเวลาจำกัดได้ เราสามารถนำข้อมูลจากสถานที่ต่าง ๆ เช่น ตลาดข้อมูล และคลังข้อมูลหรือจะกำหนดข้อมูลเอง โดยกำหนดเป้าหมายเฉพาะที่ต้องการก็ได้ ต่อจากนั้นก็ถึง เวลาทำความเข้าใจข้อมูล โดยนำข้อมูลที่ไม่เกี่ยวข้องออก

2) ขั้นตรวจวัดผล (Results Validation) หลังจากนำชุดข้อมูลออกมาแล้วจะ เป็นการตรวจวัดผล ข้อมูลที่พบอาจใช้ไม่ได้ทุกรูปแบบเสมอไป เนื่องจากการพบข้อมูลใน แบบจำลองที่ไม่เคยเห็นมาก่อน จึงต้องทำตัวอย่างข้อมูลมาทดลองและดูว่าใช้ได้จริงหรือไม่

3) ขั้นดำเนินการทำเหมืองข้อมูลมีขั้นตอนดังนี้

3.1) การหาความเบี่ยงเบน (Deviation Detection) คือการหาความผิดปกติใน ข้อมูลที่อาจใช้ได้หรืออาจจะเป็นข้อผิดพลาดที่ต้องมีการตรวจสอบ

3.2) การจัดกลุ่ม (Clustering) คือการนำข้อมูลโครงสร้างคล้าย ๆ กัน โดยไม่ ใช้ข้อมูลที่มีในโปรแกรมอยู่แล้ว

3.3) แบบจำลองความสัมพันธ์ (Dependency Modeling) คือการหา ความสัมพันธ์ของตัวแปรเช่น ร้านสะดวกซื้อรวบรวมข้อมูลจากการซื้อสินค้าของลูกค้าการใช้ แบบจำลองความสัมพันธ์จะทำให้ทางร้านสะดวกซื้อรู้ว่าลูกค้าซื้อสินค้าชนิดใดด้วยกันมากที่สุด และใช้ข้อมูลนั้นให้เป็นประโยชน์

3.4) การจำแนกข้อมูล (Classification) คือการนำข้อมูลมาจัดหมู่ใหญ่ ๆ เพื่อ สร้างข้อมูล

3.5) การวิเคราะห์การถดถอย (Regression) คือการหาฟังก์ชันที่แบบจำลองมี ข้อผิดพลาดน้อยที่สุด

### 2.1.6 การวิเคราะห์การถดถอย (Regression Analysis)

เป็นการศึกษาเกี่ยวกับความสัมพันธ์ของตัวแปรซึ่งวัตถุประสงค์หลักของการวิเคราะห์การ ถดถอยคือ ต้องการประมาณค่าของตัวแปรตัวหนึ่งเรียกว่า ตัวแปรตาม นิยมเขียนแทนด้วย Y โดย

อาศัยความรู้จากตัวแปรอื่น เรียกว่าตัวแปรอิสระนิยมนิยามแทนด้วย X หรือกล่าวอีกอย่างหนึ่งว่าเป็นการใช้ความรู้ หรือสารสนเทศจาก X เป็นเกณฑ์ในการประมาณ Y ถ้าใช้ตัวแปร X เพียงตัวแปรเดียวในการประมาณ Y และความสัมพันธ์ของ Y และ X เป็นเชิงเส้นตรง เราเรียกว่า การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression Analysis)

### 2.1.7 การวิเคราะห์สหสัมพันธ์ (Correlation Analysis)

เป็นการศึกษาระดับหรือขนาดของความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรสองตัวแปรว่ามีมากน้อยเพียงใด เครื่องมือที่ใช้วัดเรียกว่า สัมประสิทธิ์สหสัมพันธ์ นิยมนิยามแทนด้วย R โดยวัดออกมาเป็นตัวเลขที่มีค่าอยู่ระหว่าง -1 กับ 1 ถ้า R มีค่าใกล้ 1 แสดงว่า ตัวแปรสองตัวนั้นมีความสัมพันธ์กันมากและมีทิศทางเดียวกัน กล่าวคือ ถ้า X มีค่ามาก Y จะมีค่ามาก ถ้า R มีค่าใกล้ -1 แสดงว่า ตัวแปรสองตัวนั้นมีความสัมพันธ์กันมากเช่นกันแต่มีทิศทางตรงข้ามกัน กล่าวคือ ถ้า X มีค่ามาก Y จะมีค่าน้อย หรือ X มีค่าน้อย Y จะมีค่ามาก ถ้า X และ Y มีความสัมพันธ์กันน้อย ค่าสัมประสิทธิ์สหสัมพันธ์ R จะเข้าใกล้ 0

### 2.1.8 การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Linear Regression)

เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไป ซึ่งได้แก่ตัวประมาณการ (Predictor, X) และตัวตอบสนอง (Response, Y) โดยเป็นความสัมพันธ์แบบเชิงเส้น ทั้งนี้ในขั้นตอนการทำวิเคราะห์การถดถอยต้องมีการเก็บจำนวนตัวอย่างมากพอ นั่นคือ มี X และ Y ที่มีความสัมพันธ์กันหลายๆ ครั้ง เพื่อนำมาหาสมการความสัมพันธ์

### 2.1.9 การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression)

เป็นเทคนิคการวิเคราะห์สถิติ เชิงคุณภาพของข้อมูลที่ตัวแปรตามเป็นตัวแปรเชิงคุณภาพ (Qualitative Data) ซึ่งก็คือ เป็นตัวแปรเชิงกลุ่ม การวิเคราะห์การถดถอยโลจิสติกแบ่งเป็น 3 ประเภท คือ 1) การวิเคราะห์การถดถอยโลจิสติกแบบไบนารี (Binary Logistic Regression Analysis) คือการวิเคราะห์ข้อมูลโดยที่ค่าของตัวแปรตาม (Labels) มีความน่าจะเป็นเพียง 2 กลุ่ม คือมีค่าเป็น 0 กับ 1 2) การวิเคราะห์การถดถอยโลจิสติกพหุกลุ่ม (Multinomial Logistic Regression Analysis) เป็นการวิเคราะห์ข้อมูลโดยที่ค่าของตัวแปรตาม มีความน่าจะเป็นมากกว่า 2 กลุ่ม ซึ่งเป็นค่าที่ไม่สามารถ

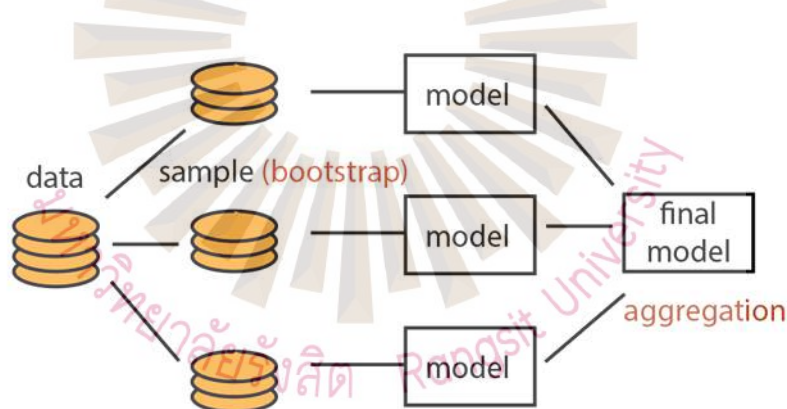


จัดลำดับได้ 3) การวิเคราะห์การถดถอยโลจิสติกเชิงลำดับ (Ordinal Logistic Regression) เป็นการวิเคราะห์ข้อมูลโดยที่ค่าของตัวแปรตาม มีความน่าจะเป็นมากกว่า 2 กลุ่ม ซึ่งเป็นค่าที่สามารถจัดลำดับได้

### 2.1.10 Ensemble Learning

เป็นการนำแบบจำลองมาเรียนรู้หลาย ๆ ครั้ง เพื่อเพิ่มประสิทธิภาพของแบบจำลองซึ่งเทคนิคที่ใช้กันบ่อยได้แก่ Bagging และ Boosting มีรายละเอียดดังนี้

2.1.10.1 Bagging หรือ Bootstrap Aggregation เป็นการสร้างแบบจำลองออกมาหลาย ๆ แบบจำลองโดยใช้การสุ่มข้อมูลตัวอย่างจากข้อมูลฝึกสอนออกมาเป็นหลาย ๆ ชุด สำหรับวิธีการสุ่มข้อมูลออกมาเป็นวิธีการสุ่มแบบแทนที่ (Random with Replacement) หมายความว่าข้อมูลที่มียังคงเดิมไม่ได้ลดลงหลังจากทำการสุ่ม ซึ่งสามารถสุ่มข้อมูลหลาย ๆ รอบเพื่อให้ได้แบบจำลองหรือตัวจำแนกข้อมูล (Classifier) หลาย ๆ ตัว เวลาทำนายจะใช้ตัวจำแนกข้อมูลทุกตัวที่สร้างขึ้นมาเพื่อทำนายชุดข้อมูลใหม่ que พบ ลักษณะการทำงานของ Bagging แสดงดังรูปที่ 2.3



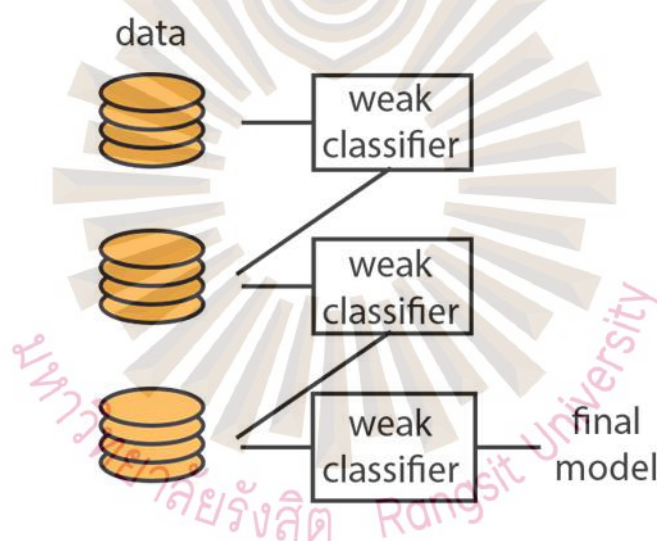
รูปที่ 2.3 แสดงลักษณะการทำงานของ Bagging

ที่มา: Achakulvisut, 2018

Bagging เป็นพื้นฐานของตัวจำแนกข้อมูลวิธีการสุ่มป่าไม้ (Random Forest Classifier) ซึ่งเป็นแบบจำลองที่นำต้นไม้ตัดสินใจ (Decision Tree) หลาย ๆ ต้นมาฝึกสอนร่วมกัน ตั้งแต่ 10 ต้นขึ้นไปจนถึง 1000 ต้น หรือมากกว่านั้น โดยที่ต้นไม้ตัดสินใจแต่ละต้นจะได้รับข้อมูลที่เป็นเซตย่อย

(Subset) ของ ข้อมูลคุณลักษณะ และข้อมูลทั้งหมดแบบสุ่ม จากนั้นก็ให้ต้นไม้ตัดสินใจแต่ละต้นทำการทำนายและเลือกผลลัพธ์จากค่าที่ได้รับการ โหวต (Vote) มากที่สุด

2.1.10.2 Boosting เป็นการนำตัวจำแนกข้อมูลที่มีความแม่นยำต่ำ (Weak Classifier) มาทำนายข้อมูล จากนั้นจะให้ตัวจำแนกข้อมูลที่มีความแม่นยำต่ำตัวใหม่มาแก้ไข ความผิดพลาด (Error) โดยผลรวมของตัวจำแนกข้อมูลจะเกิดเป็นตัวจำแนกข้อมูลใหม่ขึ้น และจะทำแบบนี้ไปจนแบบจำลองที่ได้ไม่มีค่าคลาดเคลื่อนเกิดขึ้น ซึ่งจะเป็นแบบจำลองที่ดีที่สุดลักษณะการทำงานของ Boosting แสดงดังรูปที่ 2.4 ตัวอย่างอัลกอริทึม ได้แก่ กราเดียนบูตติ้ง (Gradient Boosting : GBM) หรือ แอดาบูสต์ (Adaboost) เป็นต้น และยังมีอัลกอริทึมที่ถูกพัฒนาในลักษณะที่คล้ายกันที่น่าสนใจ และสามารถนำมาสร้างแบบจำลอง ได้แก่ ไลต์กราเดียนบูตติ้ง (Light Gradient Boosting (Light GBM) แคตบูสต์ (CatBoost) และเอ็กซ์ทรีมกราเดียนบูตติ้ง (EXtreme Gradient Boosting: XGBoost) โดยมีรายละเอียดที่จะกล่าวในหัวข้อต่อไป



รูปที่ 2.4 แสดงลักษณะการทำงานของ Boosting

ที่มา: Achakulvisut, 2018

### 2.1.11 เอ็กซ์ทรีมกราเดียนบูตติ้ง

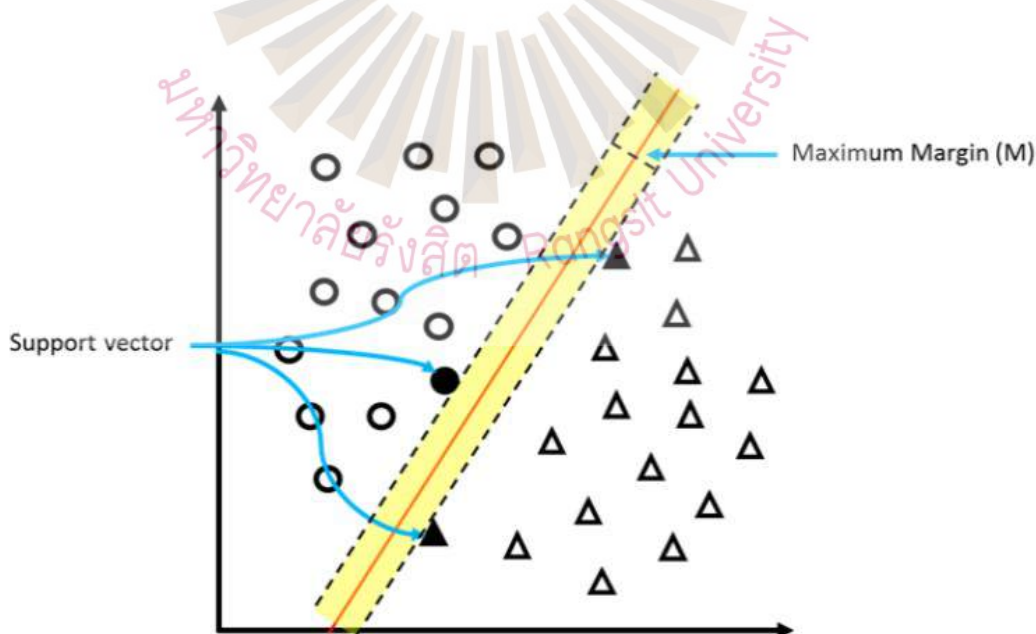
เป็นส่วนหนึ่งของกลุ่มต้นไม้ตัดสินใจซึ่งประกอบไปด้วย ต้นไม้ตัดสินใจ วิธีการสุ่มป่าไม้ Bagging, Boosting กราเดียนบูตติ้ง ดังที่กล่าวไปข้างต้นซึ่งเป็นแบบจำลองที่นำเอาแบบจำลองต้นไม้ตัดสินใจ มาฝึกสอนแบบต่อเนื่องต่อกัน โดยที่ต้นไม้ตัดสินใจแต่ละต้นฝึกจากค่าความคลาดเคลื่อน



ของต้นก่อนหน้า และเมื่อมีการฝึกของต้นไม้ตัดสินใจอย่างต่อเนื่องจนลึกมากพอ แบบจำลองก็จะหยุดฝึกเมื่อไม่มีค่าความคลาดเคลื่อนจากต้นไม้ตัดสินใจก่อนหน้าให้ฝึกแล้ว ซึ่งทำให้เกิดความแม่นยำของการทำนายมากขึ้น

### 2.1.12 วิธีการซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine Model: SVM)

เป็นอัลกอริทึมที่สามารถใช้ได้ทั้งปัญหาข้อมูลการถดถอย และปัญหาการจำแนกข้อมูล เป็นการจำแนกเชิงเส้น (Linear Classifier) แบบ 2 หมวดหมู่ โดยเกิดจากการนำค่าของกลุ่มข้อมูลมาวางลงในพื้นที่คุณลักษณะ (Feature Space) จากนั้นจึงสร้างเส้นแบ่ง (Hyperplane) ที่เป็นเส้นตรงขึ้นมาเพื่อใช้แบ่งข้อมูลทั้ง 2 หมวดหมู่ออกจากกัน และเพื่อให้ทราบว่าเส้นตรงที่แบ่ง 2 หมวดหมู่ ออกจากกันนั้น เส้นตรงใดเป็นเส้นที่ดีที่สุด โดยมีหลักการคือการหาเส้นตรงที่มี จะพิจารณาจากไฮเปอร์เพลนที่มีผลรวมของระยะห่างระหว่างเส้นแบ่งกับเส้นตรงที่ลากผ่านข้อมูลที่ใกล้ที่สุดและขนานกับเส้นแบ่งของข้อมูลแต่ละกลุ่มที่มากที่สุด (Maximum Margin) ดังตัวอย่างในภาพที่ 2.3 เป็นข้อมูลขนาด 2 มิติ ถูกจำแนกเป็น 2 หมวดหมู่ ได้แก่ + (O) และ - ( $\Delta$ ) โดยเส้นตรงที่ใช้แบ่งข้อมูลมีขอบเขต (Margin) เท่ากับ  $M=2w$  ซึ่ง เป็นความกว้างระหว่างเส้นตรงกับซัพพอร์ตเวกเตอร์ (Support Vector) ของข้อมูลทั้ง 2 หมวดหมู่

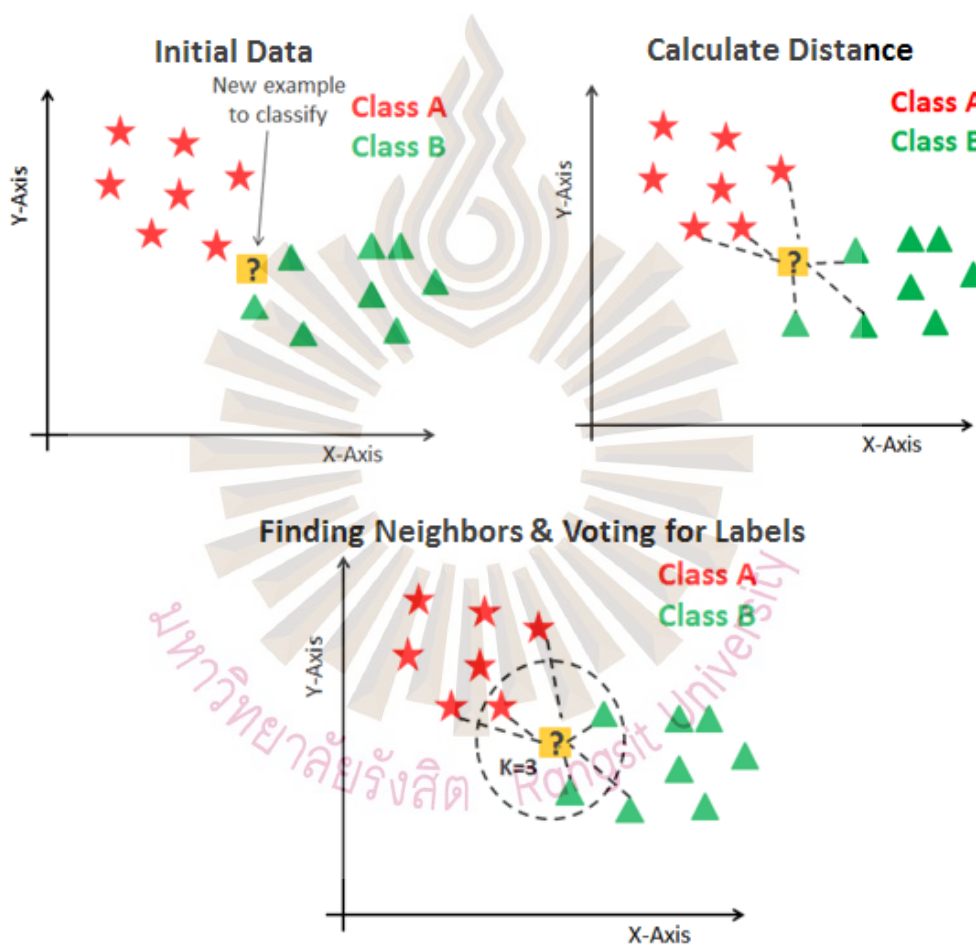


รูปที่ 2.5 แสดงวิธีการซัพพอร์ตเวกเตอร์แมชชีน

ที่มา: Navlani, 2018

### 2.1.13 วิธีการค้นหาเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor: KNN)

สามารถใช้ได้กับปัญหาข้อมูลการถดถอย และปัญหาการจำแนกข้อมูลเป็นการวิเคราะห์ข้อมูลใหม่จากข้อมูลเดิม โดยที่เราจะกำหนดค่า  $K$  เพื่อเป็นการกำหนดว่าจะหาข้อมูลที่ใกล้เคียงที่สุด (Distance) ที่ใกล้ที่สุดของข้อมูลที่ต้องการพิจารณากับกลุ่มข้อมูลตัวอย่าง โดยผลลัพธ์ของข้อมูลคือ ผลลัพธ์เดียวกับข้อมูลที่อยู่ใกล้ที่สุด  $K$  ตัว สามารถแสดงการหาเพื่อนบ้านที่ใกล้ที่สุดดังรูปที่ 2.4



รูปที่ 2.6 แสดงวิธีการค้นหาเพื่อนบ้านใกล้ที่สุด

ที่มา: Navlani, 2018

### 2.1.14 ระบบผู้เชี่ยวชาญ

การพัฒนาให้ระบบคอมพิวเตอร์มีความรู้ รู้จักใช้เหตุผลในการ วิเคราะห์ปัญหา โดยใช้ ความรู้ที่มี หรือจากประสบการณ์ในการแก้ปัญหาหนึ่ง ไปแก้ไขปัญห่อื่นอย่างมีเหตุผล ระบบนี้ จำเป็นต้องอาศัยฐานข้อมูลซึ่งมนุษย์ผู้มีความรู้ความสามารถเป็นผู้กำหนดองค์ความรู้ไว้ใน ฐานข้อมูลดังกล่าวเพื่อให้ระบบคอมพิวเตอร์สามารถวิเคราะห์ปัญหาต่าง ๆ ได้จากฐาน ความรู้นั้น จัดเป็นระบบสารสนเทศประเภทหนึ่งที่น่าวิทยาการของปัญหาประดิษฐ์เข้ามาใช้จัดการสารสนเทศ โดยเฉพาะอย่างยิ่งสารสนเทศที่เป็นองค์ความรู้ ในเฉพาะสาขาหรือเฉพาะด้าน ดังนั้นระบบ ผู้เชี่ยวชาญจึงเป็นซอฟต์แวร์ที่ใช้สร้างฐานความรู้และ กลไกในการตั้งคำถามและหาผลลัพธ์ ทำให้ ผู้ใช้ได้รับความสะดวกในการถามและตอบสิ่งที่ถามเสมือนหนึ่งคุยกับผู้เชี่ยวชาญจริง ๆ ทั้งนี้ระบบ ผู้เชี่ยวชาญจะเลือกเฉพาะสาขาหรือเฉพาะด้านที่ขาดแคลนผู้เชี่ยวชาญเท่านั้น เช่น ใช้ในงานเกี่ยวกับการวินิจฉัยทางการแพทย์ การวิเคราะห์ทางเคมี การผ่าตัด เป็นต้น อย่างไรก็ตามถึงแม้ว่าระบบ ผู้เชี่ยวชาญ จะได้รับการพัฒนา ให้ใช้งานได้สะดวก และมีประสิทธิภาพยิ่งขึ้น แต่ระบบสารสนเทศ ประเภทนี้ก็ไม่สามารถมาแทนที่มนุษย์ได้

### 2.1.15 Receiver Operating Characteristic Curve (ROC Curve)

เป็นกราฟที่ใช้วัดประสิทธิภาพของการทำนายของอัลกอริทึม ว่าสามารถจำแนก แยกแยะ ระหว่างความแตกต่างระหว่าง 2 กลุ่ม ได้แม่นยำแค่ไหน โดยที่แกน x คือ การทำนายผิดว่าเป็นผลเชิง บวก (False Positive Rate) หรือเท่ากับ  $1 - \text{Specificity}$  และ แกน Y คือ ทำนายถูกต้องว่าเป็นผลเชิง บวก (True Positive Rate) หรือเท่ากับ ค่าความไว (Sensitivity)

Confusion Matrix คือ เป็นตารางที่ใช้ประเมินผลลัพธ์การทำนาย (Predicted) เปรียบเทียบกับค่าจริง (Actual) แสดงดังตารางที่ 2.2

ตารางที่ 2.2 Confusion Matrix

		ค่าการทำนาย	
		Positive	Negative
ค่าจริง	Positive	True negative (TN)	False positive (FP)
	Negative	False negative (FN)	True positive (TP)

ที่มา: Mohajon, 2020

จากตาราง

Predicted Positive	คือ ค่าการทำนายเชิงบวก
Predicted Negative	คือ ค่าการทำนายเชิงลบ
Actually Positive	คือ ค่าจริงเชิงบวก
Actually Negative	คือ ค่าจริงเชิงลบ
TP	คือ ค่าจริงมีผลเป็นบวก และค่าทำนายมีผลเป็นบวก
TN	คือ ค่าจริงมีผลเป็นลบ และค่าทำนายมีผลเป็นลบ
FP	คือ ค่าจริงมีผลเป็นลบ และค่าทำนายมีผลเป็นบวก
FN	คือ ค่าจริงมีผลเป็นบวก และค่าทำนายมีผลเป็นลบ

แบบจำลองจะถูกนำไปวัดประสิทธิภาพ เพื่อประเมินความถูกต้องและความแม่นยำโดยสามารถประเมินได้ 4 วิธี คือ

1) ค่าความถูกต้อง (Accuracy) เป็นการวัดความถูกต้องโดยรวมของแบบจำลองระหว่างค่าจริงและค่าทำนายได้ถูกต้องใกล้เคียงกับค่าจริง โดยคำนวณได้จากสมการที่ 2-1

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2-1)$$

โดย TP, TN, FP และ FN ความหมายอ้างอิงดังตารางที่ 2.2

2) ค่าความแม่นยำ (Precision) เป็นการวัดความแม่นยำโดยวัดจากความซ้ำเดิมของแบบจำลอง ที่ทำนายได้ถูกต้องตรงกับค่าจริงโดยคำนวณได้จากสมการที่ 2-2

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

โดย TP และ FP ความหมายอ้างอิงดังตารางที่ 2.2

3) ค่าความระลึก (Recall) เป็นการวัดความครอบคลุม ซึ่งหมายถึงอัตราส่วนของการวัดค่าการทำนายได้ถูกต้องตรงกับค่าจริงจากจำนวนของค่าจริงทั้งหมด โดยครอบคลุมผลลัพธ์ที่มีทั้งหมดได้มากน้อยเพียงใด โดยคำนวณได้จากสมการที่ 2-3

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2-3)$$

โดย TP และ FN ความหมายอ้างอิงดังตารางที่ 2.2

4) ค่าความถ่วงดุล (F1-Score) คือการวัดค่าประสิทธิภาพโดยเฉลี่ยของค่า F1 เป็นการวัดค่าเฉลี่ยการทำนายระหว่าง ค่าความแม่นยำ และค่าความระลึก โดยคำนวณได้จากสมการที่ 2-4

$$F1 - Score = \left( \frac{2 \times Precision \times Recall}{Precision + Recall} \right) \quad (2-4)$$

ตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ (Area Under an ROC Curve : AUCs) คือค่าดัชนีที่ใช้บ่งชี้ความถูกต้องน่าเชื่อถือของแบบจำลอง หากแบบจำลองใดมีพื้นที่ใต้เส้นโค้ง ROC เข้าใกล้ 1 ยิ่งแสดงถึงค่าความถูกต้องการทดสอบ (Test Accuracy) มากจะถือว่าเป็นแบบจำลองที่ดีที่สุด

#### 2.1.16 ข้อมูลผู้ป่วยที่ใช้ในการวิเคราะห์

โดยทั่วไปข้อมูลผู้ป่วยจะประกอบไปด้วยรายละเอียดดังต่อไปนี้

##### 2.1.16.1 หมายเลขผู้ป่วย (Hospital Number: HN)

เป็นหมายเลขผู้ป่วยนอก ซึ่งออกหมายเลขให้ในการลงทะเบียนเป็นผู้ป่วยที่จะตรวจในโรงพยาบาล โดยอาจมีการออกหมายเลขที่ต่อเนื่อง ในบางโรงพยาบาลอาจมีการออกหมายเลขที่เป็นปีพุทธศักราชต่อท้ายหมายเลข แต่เนื่องจากหมายเลขดังกล่าวไม่นิยมให้มีการเปลี่ยนในทุกปีพุทธศักราช จึงสามารถขึ้นหมายเลขไว้ที่หน่วยเวชระเบียนผู้ป่วยนอกได้ว่าแต่ละปีมีการออกหมายเลขผู้ป่วยนอกไปตั้งแต่หมายเลขใดถึงหมายเลขใดนอกจากนี้ในหลายโรงพยาบาลยังมีการนำเลขบัตรประจำตัวประชาชน 13 หลักมาใช้ในการกำหนดหมายเลขของผู้ป่วยนอกด้วย

##### 2.1.16.2 ดัชนีมวลกาย (Body Mass Index: BMI)

ค่าดัชนีมวลกาย คือ ตัวชี้วัดมาตรฐานเพื่อประเมินสถานะของร่างกายว่ามีความสมดุลของน้ำหนักตัวต่อส่วนสูงอยู่ในเกณฑ์ที่เหมาะสมหรือไม่ ค่า BMI สามารถใช้เป็นเครื่องมือคัดกรองเพื่อระบุผู้ที่มีน้ำหนักเกิน เกณฑ์มาตรฐานในผู้ใหญ่ที่อายุ 20 ปีขึ้นไป สามารถแปลผลค่า BMI ได้ดังแสดงในตารางที่ 2.3



ตารางที่ 2.3 แสดงเกณฑ์ค่าดัชนีมวลกายในผู้ใหญ่ที่อายุ 20 ปีขึ้นไป

ค่า BMI < 18.5	แสดงถึง	อยู่ในเกณฑ์น้ำหนักน้อยหรือผอม
ค่า BMI 18.5 – 22.90	แสดงถึง	อยู่ในเกณฑ์ปกติ
ค่า BMI 23 – 24.90	แสดงถึง	น้ำหนักเกิน
ค่า BMI 25 – 29.90	แสดงถึง	โรคอ้วนระดับที่ 1
ค่า BMI 30 ขึ้นไป	แสดงถึง	โรคอ้วนระดับที่ 2

ที่มา: โรงพยาบาลบางปะกอก 9 อินเตอร์เนชั่นแนล, 2560

ในกรณีที่มีค่าดัชนีมวลกายสูง และถูกวินิจฉัยว่ามีภาวะน้ำหนักเกินหรือเป็นโรคอ้วน ก็อาจทำให้เสี่ยงต่อปัญหาสุขภาพมากมายได้แก่ โรคความดันโลหิตสูง ระดับโคเลสเตอรอลและระดับไตรกลีเซอไรด์ในเลือดสูง โรคเบาหวานชนิดที่ 2 โรคหลอดเลือดหัวใจ โรคหลอดเลือดสมอง โรคเกี่ยวกับถุงน้ำดี โรคข้อเข่าเสื่อม ภาวะการหยุดหายใจขณะหลับหรือปัญหาในการหายใจ และโรคมะเร็งชนิดต่าง ๆ

#### 2.1.16.3 ความดันสูงสุดขณะหัวใจห้องล่างบีบตัว (Systolic Blood Pressure: SBP)

ความดันสูงสุดในช่วงการหดตัวของห้องหัวใจที่ทำให้ห้องหัวใจว่างลงซึ่งเป็นผลจาก แรงดันเลือดที่ออกจากหัวใจร่วมกับการยืดขยายของหลอดเลือด

#### 2.1.16.4 ความดันต่ำสุดขณะหัวใจห้องล่างคลายตัว (Diastolic Blood Pressure: DBP)

ความดันโลหิตที่ต่ำสุดเกิดขึ้นในช่วง ระยะการคลายตัวของหัวใจเป็นผลจากแรงต้านภายในหลอดเลือดร่วมกับการปิดของลิ้นหัวใจ

#### 2.1.16.5 การตรวจระดับน้ำตาลในเลือดหลังจากอดอาหาร 8 ชั่วโมง (Fasting Blood Sugar: FBS)

การตรวจระดับน้ำตาลในเลือดหลังจากอดอาหาร 8 ชม. ใช้ในการคัดกรองและวินิจฉัยผู้ที่มีอาการแสดงหรือมีปัจจัยเสี่ยงเป็นเบาหวาน นอกจากนั้นยังใช้ติดตามระดับน้ำตาลในเลือดเพื่อประเมินผลการรักษา และตรวจป้องกันระดับน้ำตาลในเลือดสูงหรือระดับน้ำตาลในเลือดต่ำเกินไปในผู้ป่วยเบาหวาน การตรวจระดับน้ำตาลในเลือดหลังอดอาหาร 8 ชั่วโมงเป็นการบอกระดับน้ำตาลที่อยู่ในเลือด โดยปกติแล้วหลังจากการรับประทานอาหาร ร่างกายจะดูดซึมที่ลำไส้เล็กและนำไปใช้ทั่วร่างกาย โดยน้ำตาลกลูโคส เป็นแหล่งพลังงานแรกที่ร่างกายนำไปใช้ โดยเฉพาะระบบสำคัญคือ ระบบสมองและระบบประสาท ซึ่งฮอร์โมนที่เป็นตัวนำน้ำตาลไปให้ร่างกายใช้เป็นพลังงานได้นั้นคือ ฮอร์โมนอินซูลิน แต่หากร่างกายขาดฮอร์โมนอินซูลินหรือคืออินซูลินซึ่งพบ



ในผู้ป่วยเบาหวาน จะทำให้น้ำตาลอยู่ในกระแสเลือด ไม่สามารถนำไปให้อวัยวะต่าง ๆ ในร่างกายใช้ได้ ซึ่งการมีระดับน้ำตาลในกระแสเลือดสูงเป็นเวลานาน ทำให้ส่งผลต่อโรคแทรกซ้อนเช่น ไตเสื่อม หรือโรคหลอดเลือดหัวใจ เป็นต้น

#### 2.1.16.6 ฮีโมโกลบิน เอ วัน ซี (HbA1C)

การตรวจวัดค่าเฉลี่ยของน้ำตาลในเลือดที่จับกับฮีโมโกลบินของเม็ดเลือดแดง ในช่วง 3-4 เดือนที่ผ่านมาเรียกว่าน้ำตาลสะสม ซึ่งสะท้อนค่าเฉลี่ยของน้ำตาลในเลือดช่วงเวลาที่ผ่านไป เนื่องจากเม็ดเลือดแดงทั่วไปจะมีอายุขัยอยู่ที่ประมาณ 100-120 วัน ดังนั้นค่าน้ำตาลที่ตรวจได้จึงเป็นค่าน้ำตาลที่สะสมอยู่ในฮีโมโกลบินนานประมาณ 3-4 เดือน ซึ่งเป็นค่าที่ใช้พิจารณาและประเมินผลการรักษาโดยรวมในช่วงที่ผ่านมาว่าผู้ป่วยเบาหวานสามารถควบคุมระดับน้ำตาลได้ดีหรือไม่ และยังช่วยคัดกรองและวินิจฉัยได้ว่าผู้ป่วยเป็นโรคเบาหวานหรือไม่ด้วย โดยส่วนใหญ่แล้วการงดอาหาร 8 ชั่วโมงก่อนการตรวจเลือด ซึ่งค่าน้ำตาลที่ได้จะบอกได้เพียงคร่าว ๆ ว่า 1-2 วันที่ผ่านมาได้มีการรับประทานอาหารที่มีน้ำตาลมากน้อยเพียงใด ซึ่งอาจไม่เพียงพอต่อการประเมินประสิทธิภาพการรักษาในระยะยาวของแพทย์ได้ อีกทั้งบางกรณีพบว่าการงดน้ำหวานหรือขนมก่อนไปเจาะเลือด 2-3 วัน เพื่อหวังผลให้น้ำตาลในเลือดอยู่ในเกณฑ์ที่พอใจ ดังนั้นการตรวจค่าเฉลี่ยระดับน้ำตาลในเลือดในช่วง 3-4 เดือนที่ผ่านมา (HbA1c) จึงมีความสำคัญอย่างมาก เพราะช่วยประเมินผลการรักษาและป้องกันการเกิดโรคแทรกซ้อนในระยะยาวได้

#### 2.1.16.7 การตรวจหาอัตราการกรองของไต

เป็นการตรวจหาค่าอัตราการไหลของเลือดผ่านตัวกรองไตในหนึ่งนาที โดยเป็นค่าที่ได้จากการคำนวณครีเอตินิน (Creatinine) เพลส อายุ และเชื้อชาติของผู้รับการตรวจแต่ละคน ค่าครีเอตินินยิ่งสูงจะยิ่งทำให้ GFR มีค่าต่ำโดยปัจจุบันผู้เชี่ยวชาญด้านโรคไตทั่วโลก รวมทั้งสมาคมโรคไตแห่งประเทศไทย ได้เปลี่ยนมาใช้ค่า GFR ในการบอกสถานะของโรคไตแทนค่าครีเอตินิน ในคนปกติจะมีค่า GFR อยู่ที่ประมาณ 125 มล./นาที แต่ถ้าตรวจพบว่ามียาค่าต่ำกว่า 90 บ่งชี้ได้ว่าไตเริ่มเสื่อม นอกจากนี้ค่า GFR ยังสามารถบ่งชี้ความรุนแรงของโรคไตเรื้อรังว่าอยู่ในระยะใด ซึ่งระยะของโรคไตเรื้อรังแบ่งออกเป็น 5 ระยะ แสดงดังตาราง 2.4

ตารางที่ 2.4 ตารางอัตราการกรองของไต

G1	Normal or high	$\geq 90$ ml/min/1.73 m <sup>2</sup>
G2	Mildly decreased	60 - 90 ml/min/1.73 m <sup>2</sup>
G3a	Mildly to moderately decreased	45 - 59 ml/min/1.73 m <sup>2</sup>
G3b	Moderately to severely decreased	30 - 44 ml/min/1.73 m <sup>2</sup>

ตารางที่ 2.4 ตารางอัตราการทำงานของไต (ต่อ)

G4	Severely decreased	15 – 29 ml/min/1.73 m <sup>2</sup>
G5	Kidney failure	< 15 ml/min/1.73 m <sup>2</sup>

ที่มา: โรงพยาบาลสถาบันโรคไตภูมิราชนครินทร์, ม.ป.ป.

สามารถอธิบายรายละเอียดเพิ่มเติมได้ดังนี้

G1 คือ ระยะที่ 1 ค่า GFR  $\geq 90$  มล./นาที (หรือ ml/min/1.73 m<sup>2</sup>) เป็นระยะที่ตรวจพบพยาธิสภาพที่ไตแล้ว (เช่น มีนิ่ว กรวยไตอักเสบ ไตบวม) แต่ไตยังทำงานปกติ

G2 คือ ระยะที่ 2 ค่า GFR = 60 - 89 มล./นาที เป็นระยะที่ไตเริ่มทำงานผิดปกติเล็กน้อย

G3 คือ ระยะที่ 3 ค่า GFR = 30 - 59 มล./นาที เป็นระยะที่ไตทำงานผิดปกติปานกลาง

G4 คือ ระยะที่ 4 ค่า GFR = 15 - 29 มล./นาที เป็นระยะที่ไตทำงานผิดปกติอย่างมาก

G5 คือ ระยะที่ 5 ค่า GFR < 15 มล./นาที เป็นระยะสุดท้ายที่ถือว่าไตสูญเสียความสามารถในการทำงานอย่างสมบูรณ์

#### 2.1.16.8 ไขมันชนิดความหนาแน่นต่ำ (Low-Density Lipoprotein: LDL)

ไขมันที่ความหนาแน่นต่ำเป็นคอเลสเตอรอลหรือไขมันชนิดไม่ดี ส่งผลเสียต่อร่างกายและต้องถูกควบคุมไม่ให้สูงเกินไป หากมีมากเกินไปจะเข้าไปสะสมที่ผนังหลอดเลือด ทำให้เกิดหลอดเลือดตีบและแข็ง เกิดภาวะโรคหลอดเลือดแดงแข็งตัวได้ ไขมันชนิดความหนาแน่นต่ำถูกสร้างขึ้นจากตับ จากนั้นจะเข้าสู่กระแสเลือดโดยอาศัยตัวนำอย่างไลโปโปรตีน หรือ LDL-c ในการตรวจวัดค่า LDL-c มี 2 วิธีดังนี้ คือการตรวจโดยตรง (Direct LDL-c) ซึ่งวิธีนี้สามารถตรวจได้โดยการเจาะเลือดไป ซึ่งจะให้ผลที่แม่นยำ และอีกวิธีหนึ่งคือการคำนวณจากค่าของคอเลสเตอรอลตัวอื่น ได้แก่การนำผลตรวจของผลรวมคอเลสเตอรอล (Total cholesterol) ไขมันชนิดความหนาแน่นสูง (High Density Lipoprotein) และไตรกลีเซอไรด์ (Triglyceride) มาเข้าสมการเพื่อหาค่าไขมันชนิดความหนาแน่นต่ำในการตรวจไขมันชนิดความหนาแน่นต่ำส่วนมากจะใช้วิธีคำนวณโดยอาศัยวิธีนี้ เนื่องจากสะดวกประหยัด และรวดเร็วซึ่งมีสูตรการคำนวณได้จากสมการที่ 2-5

		$LDL = TC - HDL - 20\% TG$	(2-5)
โดยที่	LDL	คือ ไขมันชนิดความหนาแน่นต่ำ	
	TC	คือ ผลรวมคลอเลสเตอรอล	
	HDL	คือ ไขมันชนิดความหนาแน่นสูง	
	TG	คือ ไตรกลีเซอไรด์	

#### 2.1.16.9 การตรวจไตรกลีเซอไรด์ (Triglyceride)

ไตรกลีเซอไรด์เป็นอนุภาคไขมันชนิดหนึ่งมีขนาดเล็กที่ลอยอยู่ในเลือดหลังจากถูกดูดซึมมาทางลำไส้ใหญ่ โดยปกติร่างกายสามารถสังเคราะห์กลีเซอไรด์ขึ้นได้เองจากตับ หรืออาจมาจากอาหารต่าง ๆ ที่รับประทานเข้าไปในแต่ละวัน โดยเมื่อมีการกินอาหารเข้าไปในร่างกาย ไตรกลีเซอไรด์จะถูกส่งผ่านเข้าไปในกระแสเลือด โดยอาศัยตัวช่วยอย่างไลโปโปรตีนและส่งต่อไปยังเซลล์

## 2.2 ผลงานวิจัยที่เกี่ยวข้อง

Enchouffo-Tcheugui & Kengne (2012) ได้ศึกษาแบบจำลองความเสี่ยงในการทำนายโรคไตเรื้อรัง โดยความสัมพันธ์ของโรคไตเรื้อรังพบได้บ่อยและเกี่ยวข้องกับความเสี่ยงที่เพิ่มขึ้นของโรคหัวใจและหลอดเลือดและโรคไตวายเรื้อรังระยะสุดท้าย ซึ่งสามารถป้องกันได้โดยการระบุและรักษานบุคคลที่มีความเสี่ยงแม้ว่าจะมีการระบุปัจจัยเสี่ยงสำหรับการเกิดและการลุกลามของโรคไตเรื้อรังแต่ประโยชน์สำหรับการแบ่งชั้นความเสี่ยงโรคไตเรื้อรังผ่านแบบจำลองการทำนายยังไม่ชัดเจน เราประเมินแบบจำลองแบบจำลองความเสี่ยงอย่างละเอียดเพื่อทำนายโรคไตเรื้อรังและการลุกลามของโรคและประเมินความเหมาะสมสำหรับการใช้งานทางคลินิก การพัฒนาและการประยุกต์ใช้คะแนนความเสี่ยงต่อไตทางคลินิกอยู่ในช่วงวัยเด็กอย่างไรก็ตามการแยกแยะโรคประสิทธิภาพของเครื่องมือที่มีอยู่เป็นที่ยอมรับได้ ยังคงต้องสำรวจผลการใช้แบบจำลองเหล่านี้ในทางปฏิบัติ

Xiao et al. (2019) ได้ศึกษาแบบจำลองการวิเคราะห์ข้อมูลผู้ป่วยเพื่อทำนายการเกิดโรคไตเรื้อรัง ด้วยข้อมูลพื้นฐานและข้อมูลทางคลินิก จัดลำดับข้อมูลคุณลักษณะที่ส่งผลกระทบต่อการทำนาย โดยใช้เทคนิคการเรียนรู้ของเครื่อง ประเภทการเรียนรู้แบบมีผู้สอน ซึ่งประกอบด้วย 9 อัลกอริทึมคือ วิธีวิเคราะห์การถดถอยโลจิสติก วิธีวิเคราะห์การถดถอยแบบยืดหยุ่น (Elastic Net)

วิธีการถดถอยแบบลาสโซ (Lasso Regression) การวิเคราะห์การถดถอยแบบริดจ์ (Ridge Regression)  
วิธีการซัพพอร์ตเวกเตอร์แมชชีน วิธีการสุ่มป่าไม้ เอ็กซ์ทรีมกราเดียนบูตสตีง โครงข่ายประสาทเทียม  
วิธีการค้นหาเพื่อนบ้านใกล้ที่สุด แล้วทำการสร้างแบบจำลองเพื่อทดสอบเปรียบเทียบประสิทธิภาพ  
โดยพิจารณาด้าน ค่าความไว ค่าความจำเพาะ ความถูกต้อง ค่าความแม่นยำ



## บทที่ 3

### เครื่องมือและระเบียบวิธีการวิจัย

#### 3.1 เครื่องมือที่ใช้ในการวิจัย

ระบบวิเคราะห์ข้อมูลผู้ป่วยเบาหวานเพื่อทำนายการเกิดโรคไตเรื้อรังสามารถจำแนกวัสดุและอุปกรณ์ที่ใช้เป็นส่วนประกอบในการทำงานออกเป็น 2 ส่วนดังนี้

##### 3.1.1 ภาคนาร์ดแวร์ (Hardware) ประกอบไปด้วยอุปกรณ์ ดังนี้

3.1.1.1 คอมพิวเตอร์ซึ่งมีรายละเอียดด้านประสิทธิภาพดังต่อไปนี้

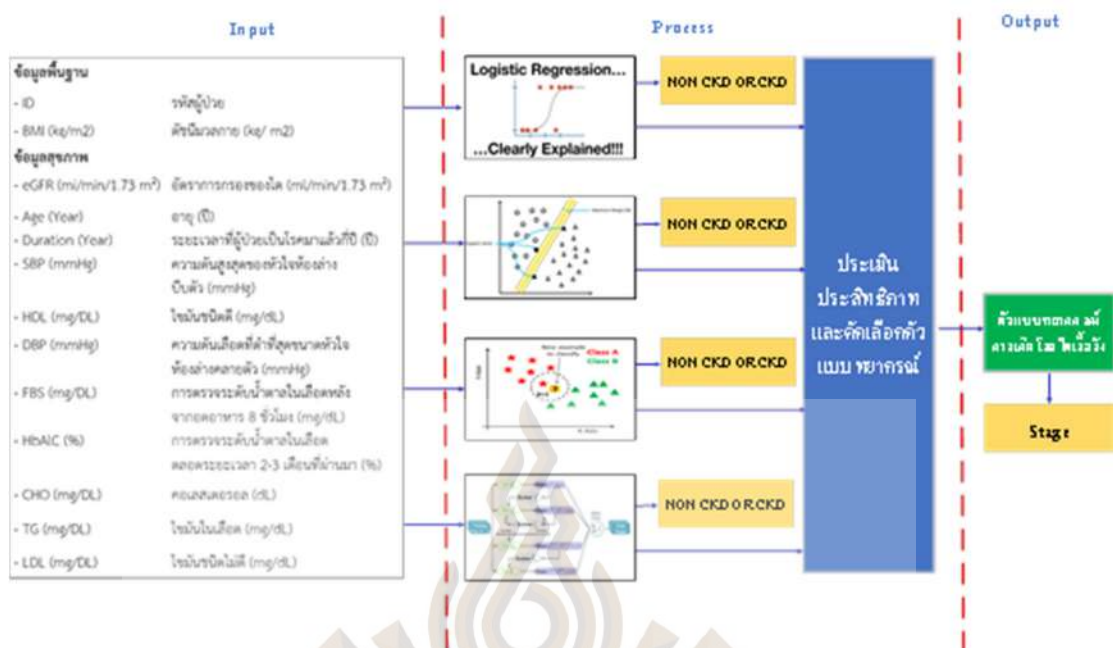
- 1) หน่วยประมวลผลกลาง: Intel Core i5 (6th Gen) 6300U / 2.4 GHz
- 2) หน่วยความจำสำรอง: 256 GB SSD M.2 SATA
- 3) หน่วยความจำหลัก: 8 GB (1 x 8 GB)
- 4) ระบบปฏิบัติการ: Windows 10 Pro 64-bit Operation System

##### 3.1.2 ภาคซอฟต์แวร์ (Software) ประกอบไปด้วยรายละเอียด ได้แก่

3.1.2.1 ภาษาที่ใช้ในการพัฒนาและเขียนระบบ ได้แก่ ภาษาไพธอน (Python)

#### 3.2 ขั้นตอนและวิธีการวิจัย

การจัดทำวิจัยการศึกษาระบบวิเคราะห์ข้อมูลผู้ป่วยเบาหวานเพื่อทำนายการเกิดโรคไตเรื้อรัง มีขั้นตอนการดำเนินงานแสดงแผนภาพแสดงดังรูปที่ 3.1



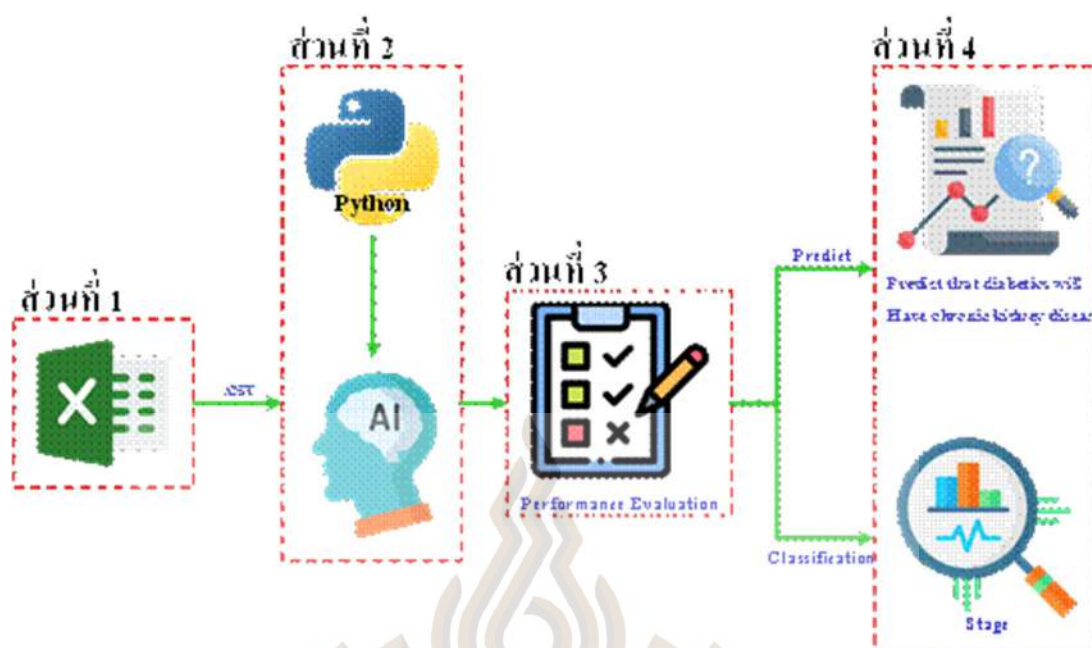
รูปที่ 3.1 แสดงขั้นตอนและวิธีการวิจัย

จากรูปที่ 3.1 แสดงการจัดทำวิจัยระบบวิเคราะห์ข้อมูลผู้ป่วยเบาหวานเพื่อทำนายการเกิดโรคไตเรื้อรัง เริ่มจากการนำข้อมูลผู้ป่วย เข้าสู่กระบวนการการวิเคราะห์เพื่อหาแบบจำลองที่เหมาะสมที่สุดและการทำนายของระบบ จากนั้นทำการทดสอบประสิทธิภาพของแบบจำลองและการประเมินความสามารถของแบบจำลองโดยเลือกแบบจำลองที่มีประสิทธิภาพสูงที่สุดเพื่อสร้างระบบที่จะแสดงผลลัพธ์ว่าในอนาคต ผู้ป่วยโรคเบาหวานเกิดโรคไตเรื้อรังหรือไม่ถ้าเกิดโรคไตเรื้อรังผู้ป่วยอยู่ในระยะใด

### 3.2.1 การออกแบบสถาปัตยกรรมของระบบ

การออกแบบสถาปัตยกรรมของระบบประกอบด้วย 4 ส่วน เริ่มจาก ส่วนที่ 1 เกี่ยวกับข้อมูล ส่วนที่ 2 เป็นการประมวลผล ส่วนที่ 3 เป็นการวิเคราะห์ และส่วนที่ 4 เป็นการแสดงผล โดยมีรายละเอียดดังแผนภาพ





รูปที่ 3.2 แสดงสถาปัตยกรรมของระบบวิเคราะห์และทำนายความเสี่ยงโรคไตเรื้อรัง

จากรูปที่ 3.2 แสดงการออกแบบสถาปัตยกรรมของระบบวิเคราะห์และทำนายความเสี่ยงโรคไตเรื้อรังของผู้ป่วยเบาหวาน ได้แบ่งวิธีการจัดทำงานวิจัยออกเป็น 4 ส่วน คือ ส่วนที่ 1 การนำข้อมูลของผู้ป่วยโรคไตเรื้อรังและ ผู้ป่วยโรคเบาหวาน ซึ่งจะประกอบด้วย ผู้ป่วยเบาหวานที่ไม่เป็นโรคไตเรื้อรังและผู้ป่วยเบาหวานที่เป็นโรคไตเรื้อรัง โดยมีข้อมูลดังนี้ 1) ข้อมูลพื้นฐานประกอบด้วยรหัสประจำตัวผู้ป่วย และค่าดัชนีมวลกาย 2) ข้อมูลสุขภาพ ประกอบด้วย ค่าความดัน ผลตรวจเลือด ผลตรวจระดับน้ำตาลการกรองของไต ผลตรวจไขมัน โดยข้อมูลทั้งหมดที่ได้มาใช้ในการเรียนรู้ของระบบ และวิเคราะห์ข้อมูล ส่วนที่ 2 พัฒนาระบบด้วยกระบวนการออกแบบระบบพื้นฐาน โดยภาษาไพธอน ส่วนที่ 3 การทดสอบเปรียบเทียบประสิทธิภาพของแบบจำลอง การประเมินความสามารถของแบบจำลอง และส่วนที่ 4 เป็นการแสดงผลการทดสอบ ได้แก่ วิเคราะห์โอกาส ที่ผู้ป่วยเบาหวานจะเป็นโรคไตเรื้อรังและวิเคราะห์ระยะของโรคไตเรื้อรัง รวมถึงระบบยัง ทำนายการเกิดโรคไตเรื้อรังในอนาคตและบอกว่าผู้ป่วยอยู่ในระยะใด

### 3.2.2 กำหนดกลุ่มจำนวนประชากร

ประชากรเป้าหมายที่ใช้ในงานวิจัยคือข้อมูลผู้ป่วยเบาหวานและข้อมูลผู้ป่วยโรคไตเรื้อรังที่เข้ารับรักษาในคณะแพทยศาสตร์วชิรพยาบาล มหาวิทยาลัยนวมินทราธิราชในระยะเวลาเก็บข้อมูล 10 ปี จำนวน 1,528 ราย

### 3.2.3 ในการเก็บและรวบรวมข้อมูลได้ปฏิบัติตามเงื่อนไขดังนี้

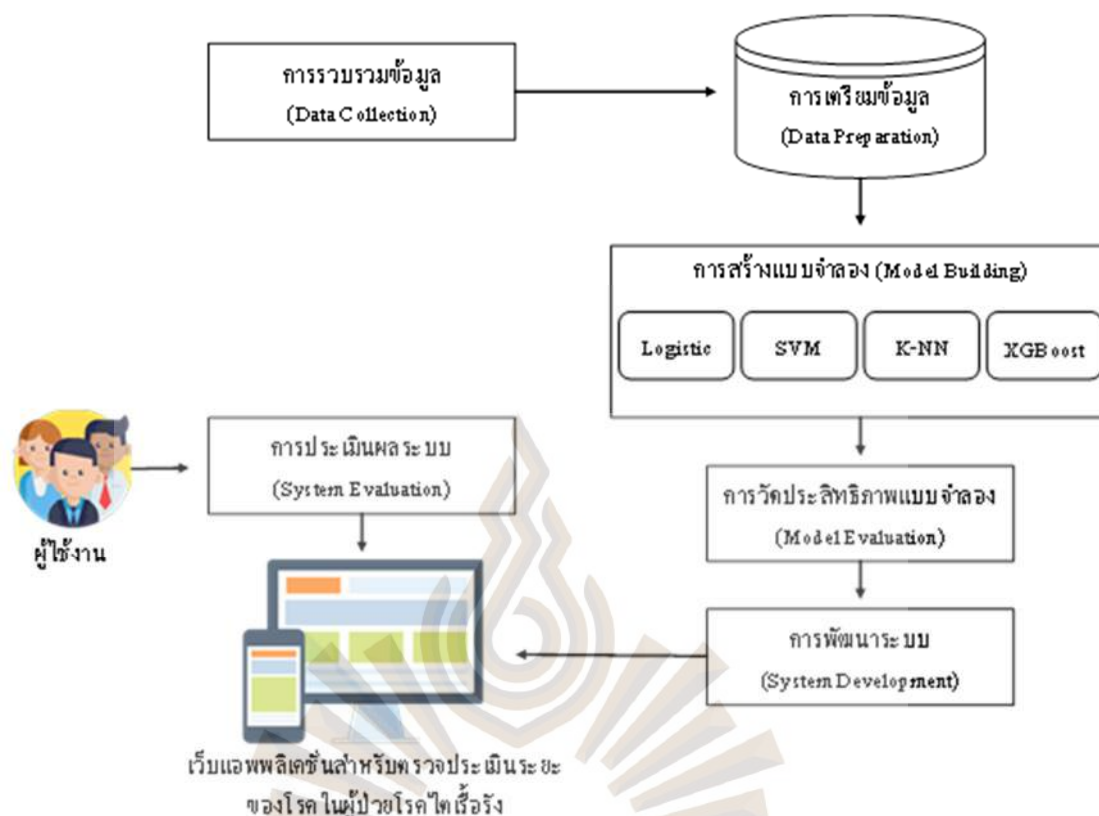
3.2.3.1 การระมัดระวังการรักษาความลับของข้อมูล เนื่องจากข้อมูลที่นำมาใช้ในการทำงานวิจัยนั้นเป็นข้อมูลของผู้ป่วยในโรงพยาบาลจึงมีความจำเป็นต้องรักษาความลับของข้อมูลเป็นอย่างดีเพื่อไม่ข้อมูลถูกนำไปใช้ในทางมิชอบ หรือทำให้เกิดความไม่สบายใจของเจ้าของข้อมูล

3.2.3.2 การเก็บข้อมูลเฉพาะส่วนที่เกี่ยวข้องกับการวิจัยเท่านั้นผู้ทำการวิจัยจะต้องเก็บหรือใช้ข้อมูลที่ตนเองจะศึกษาเท่านั้น

3.2.3.3 ข้อมูลที่ใช้ในงานวิจัยข้อมูลของผู้ป่วยโรคเบาหวาน ข้อมูลผู้ป่วยโรคไตเรื้อรัง และข้อมูลผู้ป่วยเบาหวานที่เป็นโรคไตเรื้อรัง โดยมีข้อมูล 2 ส่วน ดังนี้ ส่วนที่ 1) ข้อมูลพื้นฐาน ประกอบด้วย รหัสประจำตัวผู้ป่วย (HN) น้ำหนัก ส่วนสูง BMI หรือค่าดัชนีมวลกาย เป็นต้น และส่วนที่ 2) ข้อมูลสุขภาพ ประกอบด้วย สัญญาณชีพจร ค่าความดัน การตรวจระดับน้ำตาลในเลือดแบบ FBR การตรวจ HbA1C ตรวจระดับอัตราการกรองของไต (eGFR) ตรวจไขมันในเลือด และตรวจหาโปรตีนรั่วในปัสสาวะ

3.2.4 การสร้างระบบที่ใช้สำหรับการวิเคราะห์และทำนายผลมีขั้นตอนแสดงแผนภาพดังนี้

การสร้างระบบสำหรับการวิเคราะห์และทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน โดยมีขั้นตอนแสดงแผนภาพดังรูปที่ 3.3



รูปที่ 3.3 แสดงระบบการวิเคราะห์และทำนายการเกิดโรคไต้เรื้อรังและระยะของโรคไต้เรื้อรัง  
ในอนาคตของผู้ป่วยเบาหวาน

จากรูปที่ 3.3 แสดงกระบวนการวิเคราะห์และทำนายการเกิดโรคไต้เรื้อรังและระยะของโรคไต้เรื้อรังในอนาคตของผู้ป่วยเบาหวาน โดยแบ่งกระบวนการออกเป็น 4 ส่วน ได้แก่ การนำข้อมูลเข้าระบบ การวิเคราะห์การประมวลผล และการทำนายผล โดยมีรายละเอียดดังนี้

ส่วนที่ 1 การนำข้อมูลเข้าระบบ นำข้อมูลของผู้ป่วยโรคเบาหวานที่เป็นโรคไต้เรื้อรังจำนวน 1,528 ชุดข้อมูลในโปรแกรม Microsoft Excel กำหนดนามสกุลของไฟล์ Excel เป็น .CSV แล้วนำเข้าสู่โปรแกรม Spyder โดยใช้ภาษาไพธอน ในการเขียนอัลกอริทึม

ส่วนที่ 2 การประมวลผลกระบวนการตรวจสอบ การแก้ไข การลบข้อมูลที่ซ้ำซ้อน หรือแทนค่า ข้อมูลที่ว่างด้วยค่า Nan ที่แสดงถึงค่าว่างที่ไม่สามารถนำไปวิเคราะห์ได้เพื่อให้รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูลแล้วแทนที่ด้วยค่าเฉลี่ย ซึ่งการแทนที่ด้วยค่าเฉลี่ย จะสามารถแทนได้ในบางชุดข้อมูล เนื่องจากข้อมูลที่สามารทดแทนด้วยค่าเฉลี่ย ได้จะต้อง

เป็นข้อมูลที่ได้จากการใช้สมการคำนวณเท่านั้น โดยจะไม่สามารถแทนที่ค่ามาตรฐาน ได้และการหาความสัมพันธ์ ของข้อมูลเพื่อดูทิศทางความสัมพันธ์ระหว่างตัวแปร 2 ตัว โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เป็นตัวบ่งชี้ถึงความสัมพันธ์นี้ ซึ่งค่าสัมประสิทธิ์สหสัมพันธ์นี้จะมีค่าอยู่ ระหว่าง -1.0 ถึง +1.0 ซึ่งหากมีค่าใกล้ -1.0 นั้นหมายความว่าตัวแปรทั้งสองตัวมีความสัมพันธ์กัน อย่างมากในเชิงตรงกันข้าม หากมีค่าใกล้ +1.0 นั้นหมายความว่าตัวแปรทั้งสองมีความสัมพันธ์กัน โดยตรงอย่างมาก และหากมีค่าเป็น 0 นั้นหมายความว่าตัวแปรทั้งสองตัวไม่มีความสัมพันธ์ต่อกัน ด้วยการเลือกตัวแปรอิสระกับตัวแปรตาม ที่มีค่าสัมพันธ์ที่อยู่ในช่วง(-1,-0.3) และ (0.3,1) จากนั้น ทำการหาค่า P-Value เพื่อหากระดับนัยสำคัญ ของข้อมูลค่าสหสัมพันธ์เป็นมาตรวัดของความ แข็งแรงและทิศทางของความสัมพันธ์เชิงเส้นระหว่างสองตัวแปร มีค่าอยู่ในช่วง [-1,1] ระบบจะ ทำการประมวลผลชุดข้อมูลที่กำหนดไว้ด้วยกระบวนการทดสอบระบบ โดยแบ่งการเรียนรู้ของ ระบบออก เป็น 70% ของข้อมูลทั้งหมด และแบ่งการทดสอบของระบบออกเป็น 30 % ของข้อมูล ทั้งหมด

ส่วนที่ 3 ส่งเข้าการเรียนรู้ของเครื่อง ประเภทการเรียนรู้ของเครื่องแบบมีผู้สอน ซึ่งประกอบด้วย อัลกอริทึมการวิเคราะห์การถดถอยโลจิสติก อัลกอริทึมวิธีการซัพพอร์ตเวกเตอร์แมชชีน อัลกอริทึมวิธีการค้นหาเพื่อนบ้านใกล้ที่สุด และอัลกอริทึมเอ็กซ์ทริมกราเดียนบูตติ้ง โดยอัลกอริทึมทั้งหมดใช้ค่าพารามิเตอร์มาตรฐานมาทำการเรียนรู้ แล้วทำการทดสอบเปรียบเทียบประสิทธิภาพของแบบจำลอง การประเมินความสามารถของแบบจำลอง ใช้วิธีการทดสอบเพื่อเปรียบเทียบประสิทธิภาพของแบบจำลอง โดยพิจารณาจาก Confusion Matrix คือการประเมินผลลัพธ์การทำนาย เปรียบเทียบกับค่าจริงในรูปแบบตาราง โดยพิจารณาจากค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึกลับ ค่าความถ่วงดุล และค่าตัววัดประสิทธิภาพ โดยใช้พื้นที่ใต้กราฟ ซึ่งเป็นวิธีที่เป็นมาตรฐานในการทดสอบแบบจำลองด้านการเรียนรู้ของเครื่องประเภทการเรียนรู้แบบมีผู้สอน

ส่วนที่ 4 แบ่งเป็นสองส่วนคือการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังและการวิเคราะห์ระยะการเกิดโดยมีรายละเอียดดังนี้

1) การวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน โดยมีเกณฑ์การ วิเคราะห์ ถ้าผู้ป่วยเบาหวานมีระดับอัตราการกรองของไตมากกว่า 60 มิลลิตร/นาที/1.73ตารางเมตร ระบบจะถือผู้ป่วยเบาหวานไม่เป็น โรคไตเรื้อรัง แต่ถ้าระดับอัตราการกรองของไตต่ำกว่า 60 มิลลิตร/นาที/1.73ตารางเมตร จะถือว่าผู้ป่วยเบาหวานเป็นโรคไตเรื้อรัง

2) การวิเคราะห์ระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน โดยมีเกณฑ์การ วิเคราะห์ระยะของโรคไตเรื้อรัง ดังตารางที่ 2.1

### 3.3 ขั้นตอนและวิธีการทดสอบการวิจัย

สามารถแบ่งเป็นการทดสอบย่อยได้ดังนี้

#### 3.3.1 การทดสอบระบบสำหรับเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน

จากการทดสอบระบบสำหรับเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน เพื่อให้ได้แบบจำลองที่มีประสิทธิภาพสูงสุด

3.3.1.1 การทดสอบระบบโดยการนำเข้าสู่ข้อมูลของผู้ป่วยโรคเบาหวาน โดยข้อมูลที่นำเข้ดังตารางที่ 3.1

ตารางที่ 3.1 แสดงข้อมูลที่นำเข้ระบบสำหรับเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน

ลำดับ	ข้อมูล	ความหมาย
1	ID	รหัสผู้ป่วย
2	eGFR (ml/min/1.73 m <sup>2</sup> )	อัตราการกรองของไต (ml/min/1.73 m <sup>2</sup> )
3	Age (Year)	อายุ (ปี)
4	Duration (Year)	ระยะเวลาที่ผู้ป่วยเป็นโรคมานานกี่ปี (ปี)
5	BMI (kg/m <sup>2</sup> )	ดัชนีมวลกาย (kg/m <sup>2</sup> )
6	SBP (mmHg)	ความดันสูงสุดของหัวใจห้องล่างบีบตัว (mmHg)
7	HDL (mg/dL)	ไขมันชนิดดี (mg/dL)
8	DBP (mmHg)	ความดันเลือดที่ต่ำที่สุดขนาดหัวใจห้องล่างคลายตัว (mmHg)
9	FBS (mg/dL)	การตรวจระดับน้ำตาลในเลือดหลังจากอดอาหาร 8 ชั่วโมง (mg/dL)
10	HbA1C (%)	การตรวจระดับน้ำตาลในเลือดตลอดระยะเวลา 2 – 3 เดือนที่ผ่านมา (%)
11	CHO (mg/dL)	คอเลสเตอรอล (dL)
12	TG (mg/dL)	ไตรกลีเซอไรด์ (mg/dL)
13	LDL (mg/dL)	ไขมันชนิดความหนาแน่นต่ำ (mg/dL)



จากตารางที่ 3.1 แสดงข้อมูลที่นำเข้าระบบเพื่อใช้ในการเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานแสดงดังรูปที่ 3.4

ID	eGFR	Age	Duration	BMI	SBP1	DBP1	FB51	A1c1	CHO1	TG1	HDL1	LDL1	Stage	
1044	21202	88	65	12	23.49	139	88	104	8.4	182	88	45	119	0
1046	10139	89	73	17	29.9	157	83	130	6.8	152	114	54	75	0
1047	10149	89	70	15	29.74	128	78	155	7.3	155	142	41	86	0
1048	10206	89	60	18	26.78	126	77	183	9.6	121	89	35	68	0
1049	10233	89	70	5	28.63	150	61	144	7.6	149	199	48	61	0
1050	20052	89	56	14	35.56	145	76	130	8.6	282	121	61	197	0
1051	20237	89	57	24	22.62	155	85	213	9	172	152	48	94	0
1052	20283	89	77	8	17.94	145	78	156	10.4	204	120	43	137	0
1053	20325	89	74	21	36.12	147	62	93	5.9	190	125	61	104	0
1054	20363	89	80	21	22.27	135	68	154	7.7	128	61	51	65	0
1055	20399	89	67	24	26.57	118	75	115	7.5	116	70	41	61	0
1056	20404	89	69	15	24.06	140	71	124	8.2	132	136	27	78	0
1057	20603	89	56	10	30.93	150	101	158	10.1	207	181	36	135	0
1058	20692	89	70	15	29.42	133	70	121	6.5	175	82	62	97	0
1059	20772	89	72	9	28.27	147	75	101	6.6	154	280	29	69	0
1060	20839	89	83	16	32.39	163	92	155	6.9	162	151	33	99	0
1061	21015	89	50	9	23.48	111	72	148	7	248	87	57	174	0
1062	21120	89	77	47	20.82	134	76	91	6.8	204	84	77	110	0
1063	21127	89	62	6	20.49	90	40	127	6.7	189	94	44	126	0
1064	21166	89	83	13	29.2	158	69	96	6.1	227	81	60	151	0
1065	10013	20	76	13	22.74	171	76	166	12.2	219	292	56	105	1
1066	10019	16	77	22	24.19	153	81	138	7	121	302	28	111	1
1067	10073	15	63	17	29.73	181	85	181	8.1	276	152	72	174	1
1068	10107	28	90	12	29.78	172	92	85	7.1	134	65	44	77	1
1069	10169	23	68	23	23.32	161	80	84	7.6	176	169	56	86	1
1070	10234	21	88	23	19.98	100	70	234	11.1	157	62	59	86	1
1071	10244	24	64	19	23.15	153	91	195	8.9	270	61	72	186	1
1072	10255	22	80	25	21.75	135	85	166	9.1	250	81	68	166	1
1073	10296	29	74	20	26.12	121	80	340	10.1	202	188	28	136	1
1074	10298	27	89	26	27.92	123	63	94	8.6	167	124	37	105	1
1075	10345	21	76	21	27.64	140	74	134	10.6	113	163	24	56	1
1076	10347	16	49	13	29.71	122	83	245	9.8	261	269	57	150	1
1077	10352	21	61	9	30.74	146	91	260	13.5	146	110	40	84	1
1078	20012	26	76	21	21.64	125	67	281	10.5	118	145	32	57	1
1079	20025	26	69	9	18.18	153	92	280	13.8	185	120	67	94	1

รูปที่ 3.4 แสดงการนำข้อมูลเข้าระบบ

จากรูปที่ 3.4 เมื่อกำหนดนามสกุลของไฟล์ Excel เป็น .csv แล้วนำข้อมูลของผู้ป่วยโรคเบาหวานเข้าสู่โปรแกรม Spyder โดยใช้ภาษาไพธอน ในการเขียนอัลกอริทึม

3.3.1.2 การแทนค่าที่ไม่สมบูรณ์หรือข้อมูลที่ขาดหายไปด้วยค่าเฉลี่ย ในภาษาไพธอนนั้น สามารถหาค่าเฉลี่ยได้โดยใช้คำสั่ง np.mean() ซึ่งจะหาค่าเฉลี่ยของข้อมูลและสามารถหาค่าที่ไม่สมบูรณ์หรือข้อมูลที่ขาดหายไปด้วยคำสั่ง .fillna() และแทนที่ข้อมูลที่ขาดหายไปด้วยค่าเฉลี่ย

3.3.1.3 การทดสอบระบบ ระบบจะทำการประมวลผลชุดข้อมูลที่กำหนดไว้ด้วยกระบวนการทดสอบระบบ โดยแบ่งการเรียนรู้ของระบบออกเป็น 70% ของข้อมูลทั้งหมด และแบ่งการทดสอบของระบบออกเป็น 30 % ของข้อมูลทั้งหมด

3.3.1.4 การทดสอบระบบ ระบบสร้างแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน โดยใช้อัลกอริทึมดังนี้

- 1) อัลกอริทึมการวิเคราะห์การถดถอยโลจิสติก
- 2) อัลกอริทึมวิธีการค้นหาเพื่อนบ้าน ใกล้ที่สุด
- 3) อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน



ดังนี้

4) อัลกอริทึมเอ็กซ์ทรีมกราเดียนบูตดิ้ง

3.3.1.5 การทดสอบระบบ โดยใช้เครื่องมือในการวัดประสิทธิภาพของแบบจำลอง

1) Confusion matrix

2) ค่าความถูกต้อง

3) ค่าความแม่นยำ

4) ค่าความระลึก

5) ค่าความถ่วงดุล

6) ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ

### 3.3.2 การทดสอบระบบสำหรับการวิเคราะห์ระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน

จากการทดสอบระบบวิเคราะห์ข้อมูลผู้ป่วยเบาหวานเพื่อทำนายการเกิดโรคไตเรื้อรัง โดยใช้อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน เพื่อวิเคราะห์ระยะของโรคไตเรื้อรังในผู้ป่วยเบาหวาน

3.3.2.1 การทดสอบระบบโดยการนำเข้าข้อมูลของผู้ป่วยโรคเบาหวาน โดยข้อมูลที่นำเข้ดังตารางที่ 3.1

3.3.2.2 การทดสอบการหาความสัมพันธ์ของข้อมูล คุณลักษณะความสัมพันธ์ระหว่างตัวแปร 2 ตัว โดยมี ค่าสัมประสิทธิ์สหสัมพันธ์ เป็นตัวบ่งชี้ถึงความสัมพันธ์นี้ ซึ่งค่าสัมประสิทธิ์สหสัมพันธ์นี้จะมีค่าอยู่ระหว่าง -1.0 ถึง +1.0 ซึ่งหากมีค่าใกล้ -1.0 นั้นหมายความว่าตัวแปรทั้งสองตัวมีความสัมพันธ์กันอย่างมากในเชิงตรงกันข้าม หากมีค่าใกล้ +1.0 นั้นหมายความว่า ตัวแปรทั้งสองมีความสัมพันธ์กันโดยตรงอย่างมาก และหากมีค่าเป็น 0 นั้นหมายความว่า ตัวแปรทั้งสองตัวไม่มีความสัมพันธ์ต่อกัน นั้นทำการหานัยสำคัญทางสถิติของข้อมูล เพื่อทดสอบสมมติฐาน

3.3.2.3 การทดสอบการแทนค่าข้อมูลลงในระบบกำหนดตัวแปรอิสระ และกำหนดตัวแปรตามโดยตั้งสมมติฐานทางสถิติดังนี้

1) สมมติฐานเชิงสถิติของระดับอัตราการกรองของไต ส่งผลต่อระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานหรือไม่ โดยมีเงื่อนไขดังนี้คือ

$H_0$  : ระดับอัตราการกรองของไต ไม่ส่งผลต่อระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$

$H_a$  : ระดับอัตราการกรองของไต ส่งผลต่อระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_1 \neq 0$

2) สมมติฐานเชิงสถิติ อายุ ส่งผลต่อผู้ป่วยเบาหวานที่เป็นโรคไตเรื้อรังและไม่เป็นโรคไตเรื้อรังหรือไม่

$H_0$  : อายุ ไม่ส่งผลต่อผู้ป่วยเบาหวานที่เป็นโรคไตเรื้อรังและไม่เป็นโรคไตเรื้อรัง (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$

$H_a$  : อายุ ส่งผลต่อผู้ป่วยเบาหวานที่เป็นโรคไตเรื้อรังและไม่เป็นโรคไตเรื้อรัง (สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_2 \neq 0$

3) สมมติฐานเชิงสถิติของระยะเวลาที่ผู้ป่วยเป็นโรค (ปี) ส่งผลต่อระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานหรือไม่

$H_0$  : ระยะเวลาที่ผู้ป่วยเป็นโรค (ปี) ไม่ส่งผลต่อระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$

$H_a$  : ระยะเวลาที่ผู้ป่วยเป็นโรค (ปี) ส่งผลต่อระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_3 \neq 0$

4) สมมติฐานเชิงสถิติของดัชนีมวลกาย ส่งผลต่อระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานหรือไม่

$H_0$  : ดัชนีมวลกาย ไม่ส่งผลต่อระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$

$H_a$  : ดัชนีมวลกาย ส่งผลต่อระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_4 \neq 0$

5) สมมติฐานเชิงสถิติของความดันสูงสุดของหัวใจห้องล่างบีบตัว ส่งผลต่อระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานหรือไม่

$H_0$  : ความดันสูงสุดของหัวใจห้องล่างบีบตัว ไม่ส่งผลต่อระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$

$H_a$  : ความดันสูงสุดของหัวใจห้องล่างบีบตัว ส่งผลต่อระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_5 \neq 0$

6) สมมติฐานเชิงสถิติของไขมันชนิดดี ส่งผลต่อระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานหรือไม่

$H_0$  : ไขมันชนิดดี ไม่ส่งผลต่อระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$

$H_a$  : ไขมันชนิดดี ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_6 \neq 0$

7) สมมติฐานเชิงสถิติของความดันเลือดที่ต่ำที่สุดขนาดหัวใจห้องล่างคลายตัว ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานหรือไม่

$H_0$  : ความดันเลือดที่ต่ำที่สุดขนาดหัวใจห้องล่างคลายตัว ไม่ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$

$H_a$  : ความดันเลือดที่ต่ำที่สุดขนาดหัวใจห้องล่างคลายตัว ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_7 \neq 0$

8) สมมติฐานเชิงสถิติของการตรวจระดับน้ำตาลในเลือดหลังจากอดอาหาร 8 ชั่วโมง ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานหรือไม่

$H_0$  : การตรวจระดับน้ำตาลในเลือดหลังจากอดอาหาร 8 ชั่วโมง ไม่ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$

$H_a$  : การตรวจระดับน้ำตาลในเลือดหลังจากอดอาหาร 8 ชั่วโมง ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_8 \neq 0$

9) สมมติฐานเชิงสถิติของการตรวจระดับน้ำตาลในเลือดตลอดระยะเวลา 2 – 3 เดือนที่ผ่านมา (HbA1c) ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานหรือไม่

$H_0$  : การตรวจระดับน้ำตาลในเลือดตลอดระยะเวลา 2 – 3 เดือนที่ผ่านมา ไม่ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$

$H_a$  : การตรวจระดับน้ำตาลในเลือดตลอดระยะเวลา 2 – 3 เดือนที่ผ่านมา ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_9 \neq 0$

10) สมมติฐานเชิงสถิติของคลอเลสเทอรอล ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานหรือไม่

$H_0$  : คลอเลสเทอรอล ไม่ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$

$H_a$  : คลอเลสเทอรอล ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_{10} \neq 0$

11) สมมติฐานเชิงสถิติของไตรกลีเซอไรด์ ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานหรือไม่

$H_0$  : ไตรกลีเซอไรด์ ไม่ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$

$H_a$  : ไตรกลีเซอไรด์ ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_{11} \neq 0$

12) สมมติฐานเชิงสถิติของไขมันชนิดความหนาแน่นต่ำ ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานหรือไม่

$H_0$  : ไขมันชนิดความหนาแน่นต่ำ ไม่ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$

$H_a$  : ไขมันชนิดความหนาแน่นต่ำ ส่งผลต่อระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน (สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_{12} \neq 0$

3.3.2.4 การแทนค่าที่ไม่สมบูรณ์หรือข้อมูลที่ขาดหายไปด้วยค่าเฉลี่ย ในภาษาไพธอนนั้น สามารถหาค่าเฉลี่ยได้โดยใช้คำสั่ง `np.mean()` ซึ่งจะได้ค่าเฉลี่ยของข้อมูลและสามารถหาค่าที่ไม่สมบูรณ์หรือข้อมูลที่ขาดหายไปด้วยคำสั่ง `.fillna()` และแทนที่ข้อมูลที่ขาดหายไปด้วยค่าเฉลี่ย

3.3.2.5 การทดสอบระบบ ระบบจะทำการประมวลผลชุดข้อมูลที่กำหนดไว้ด้วยกระบวนการทดสอบระบบ โดยแบ่งการเรียนรู้ของระบบออกเป็น 70% ของข้อมูลทั้งหมด และแบ่งการทดสอบของระบบออกเป็น 30 % ของข้อมูลทั้งหมด

3.3.2.6 การทดสอบระบบโดยใช้เครื่องมือในการวัดประสิทธิภาพของแบบจำลอง

ดังนี้

1) Confusion matrix

2) ค่าความถูกต้อง

3) ค่าความแม่นยำ

4) ค่าความระลึก

5) ค่าความถ่วงดุล

6) ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ

3.3.2.7 การทดสอบการวิเคราะห์ระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน ระบบจะทำการวิเคราะห์ระยะเวลาการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน โดยมีเกณฑ์การวิเคราะห์ระยะของโรคไตเรื้อรัง ดังตารางที่ 2.1

### 3.3.3 การทดสอบระบบสำหรับการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรัง ในอนาคตของผู้ป่วยเบาหวาน

จากการทดสอบระบบวิเคราะห์ข้อมูลผู้ป่วยเบาหวานเพื่อทำนายการเกิดโรคไตเรื้อรัง โดยใช้อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน เพื่อทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน

3.3.3.1 การทดสอบระบบโดยการนำเข้าข้อมูลของผู้ป่วยโรคเบาหวาน โดยข้อมูลที่นำเข้าดังตารางที่ 3.2

ตารางที่ 3.1 แสดงข้อมูลที่นำเข้าระบบสำหรับการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน

ลำดับ	ข้อมูล	ความหมาย
1	ID	รหัสผู้ป่วย
2	Vstdate_first	วันแรกของการเข้ารักษาที่โรงพยาบาล
3	Vstime_first	วันที่ตรวจพบระดับค่า eGFR ครั้งแรก
4	eGFR1 (ml/min/1.73 m <sup>2</sup> )	อัตราการกรองของไตครั้งแรก (ml/min/1.73 m <sup>2</sup> )
5	Vstime_change	วันที่ตรวจพบระดับค่า eGFR เปลี่ยนไป
6	eGFR2 (ml/min/1.73 m <sup>2</sup> )	อัตราการกรองของไตครั้งที่เปลี่ยนไป (ml/min/1.73 m <sup>2</sup> )
7	Duration (Day)	ระยะเวลาการเกิดโรคไตเรื้อรังในผู้ป่วยเบาหวาน
8	Age (Year)	อายุ (ปี)
9	Duration (Year)	ระยะเวลาที่ผู้ป่วยเบาหวานเป็นโรคมานานกี่ปี (ปี)
10	BMI_change (kg/m <sup>2</sup> )	ดัชนีมวลกายที่เปลี่ยนไป(kg/ m <sup>2</sup> )
11	SBP_change (mmHg)	ความดันสูงสุดของหัวใจห้องล่างบีบตัวที่เปลี่ยนไป (mmHg)
12	HDL_change (mg/dL)	ไขมันชนิดดีที่เปลี่ยนไป(mg/dL)
13	DBP_change (mmHg)	ความดันเลือดที่ต่ำที่สุดขนาดหัวใจห้องล่างคลายตัวที่เปลี่ยนไป (mmHg)



ตารางที่ 3.2 แสดงข้อมูลที่น่าใช้ระบบสำหรับการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน (ต่อ)

ลำดับ	ข้อมูล	ความหมาย
14	FBS_change (mg/dL)	การตรวจระดับน้ำตาลในเลือดหลังจากอดอาหาร 8 ชั่วโมงที่เปลี่ยนไป (mg/dL)
15	HbA1C_change (%)	การตรวจระดับน้ำตาลในเลือดตลอดระยะเวลา 2 – 3 เดือนที่ผ่านมา (%)
16	CHO_change (mg/dL)	คอเลสเตอรอลที่เปลี่ยนไป (dL)
17	TG_change (mg/dL)	ไขมันในเลือดที่เปลี่ยนไป (mg/dL)
18	LDL_change (mg/dL)	ไขมันชนิดไม่ดีที่เปลี่ยนไป (mg/dL)

เมื่อนำเข้าข้อมูลในตารางที่ 3.2 เข้าสู่ระบบสำหรับการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน โดยใช้ข้อมูลของผู้ป่วยโรคเบาหวาน แสดงลักษณะการนำเข้าข้อมูลได้ดังรูปที่ 3.2

ID	Vstartdate	Vsttime	Vfirst	eGFR1	Vsttime	char	eGFR2	duration	(day)	Age	Duration	BMI_change	SBP_change	DBP_change	FBS_change	A1c_change	CHO_change	TG_change	HDL_change	LDL_change	Stage	stated_change
1	10002	9/5/2006	16/10/2012	42.7	25/3/2015	53	1260	76	19	21.05	161	98	167	6.9	182	235	28	107	3	-1		
2	10004	9/7/2007	11/12/2012	24.2	1/1/2016	11	1116	81	39	27.06	160	89	122	7.7	151	179	36	79	5	1		
3	10006	25/12/2009	10/10/2014	55	26/3/2015	60	167	79	18		137	102	83							3	1	
4	10006	26/3/2015	60	1/1/2019	44	1377	83	22	34.6	205	95	88	6	109	99	36	53	3	-1			
5	10007	19/4/2016	30	1/1/2019	55	987	67	14	29.62	136	76	172	8.3	179	128	43	110	3	-1			
6	10007	27/7/2006	10/7/2012	53.7	12/2/2013	44.3	217	64	11		98	76	172	8.3	179	128	43	110	3	1		
7	10007	27/5/2014	83	9/12/2014	56	196	65	12		124	76	174	6	148	215	29	3	1				
8	10019	7/4/2016	22	1/1/2019	16	999	77	22	24.19	153	81	138	7	121	302	28	111	4	1			
9	10019	16/4/2015	39	7/4/2016	22	357	75	20	23.53	159	85	112	7.3	129	198	35	72	4	-1			
10	10019	26/12/2013	26/12/2013	51	30/1/2014	79	35	74	19		204	114	248	11.6	157	178	25	2	-1			
11	10020	23/1/2013	6/6/2014	50	17/9/2015	53	468	74	17	25.96	136	60	148	8.4	185	212	46	109	3	-1		
12	10020	7/1/2016	51	1/1/2019	48	1090	77	17	25.96	141	81	259	9.3	186	144	43	114	3	1			
13	10022	7/1/2016	9	1/1/2019	9	1090	63	26	29.78	111	84	114	7.5	185	72	37	134	5	1			
14	10022	23/9/2014	10	30/6/2015	8	180	59	26		128	56	309	7	309	90	48	151	5	1			
15	10022	1/2/2005	20/6/2011	15.5	11/12/2012	26	540	58	26		117	84	119	10.7	158	142	35	5	-1			
16	10029	2/12/2013	2/12/2013	40	8/9/2015	38	645	63	2		133	71	188	9.1	197	150	57	4	1			
17	10029	4/4/2016	46	1/1/2019	42	1002	84	7	34.13	147	90	195	8.5	165	192	59	68	3	1			
18	10032	23/5/2016	81	1/12/2016	59	192	63	27	30.51	119	70	210	10.1	260	184	36	187	2	1			

รูปที่ 3.5 แสดงข้อมูลที่ใช้ในการทดสอบระบบ

จากรูปที่ 3.2 ข้อมูลที่ใช้ในการทดสอบระบบเมื่อกำหนดนามสกุลของไฟล์ Excel เป็น .csv แล้วนำข้อมูลของผู้ป่วยโรคเบาหวานเข้าสู่โปรแกรม Spyder โดยใช้ภาษาไพธอน ในการเขียนอัลกอริทึม

3.3.3.2 การทดสอบการหาความสัมพันธ์ของข้อมูล คุณลักษณะความสัมพันธ์ระหว่างตัวแปร 2 ตัว โดยมีค่าสัมประสิทธิ์สหสัมพันธ์ เป็นตัวบ่งชี้ถึงความสัมพันธ์นี้ นั้นทำการหาข้อมูลสำคัญทางสถิติของข้อมูล เพื่อทดสอบสมมติฐาน



3.3.3.3 การทดสอบการแทนค่าข้อมูลลงในระบบกำหนดตัวแปรอิสระ และกำหนดตัวแปรตามโดยตั้งสมมติฐานทางสถิติดังนี้

1) สมมติฐานเชิงสถิติของระดับอัตราการกรองของไตครั้งแรก ส่งผลต่อการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวานหรือไม่

$H_0$  : ระดับอัตราการกรองของไตครั้งแรก ไม่ส่งผลต่อการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$

$H_a$  : ระดับอัตราการกรองของไตครั้งแรก ส่งผลต่อการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน (สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_1 \neq 0$

2) สมมติฐานเชิงสถิติของอายุ ส่งผลต่อการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวานหรือไม่

$H_0$  : อายุ ไม่ส่งผลต่อการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$

$H_a$  : อายุ ส่งผลต่อการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน (สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_2 \neq 0$

3) สมมติฐานเชิงสถิติของระยะเวลาที่ผู้ป่วยเบาหวานเป็นโรค (ปี) ส่งผลต่อการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวานหรือไม่

$H_0$  : ระยะเวลาที่ผู้ป่วยเบาหวานเป็นโรค (ปี) ไม่ส่งผลต่อการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$

$H_a$  : ระยะเวลาที่ผู้ป่วยเบาหวานเป็นโรค (ปี) ส่งผลต่อการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน (สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_3 \neq 0$

4) สมมติฐานเชิงสถิติของดัชนีมวลกายที่เปลี่ยนไป ส่งผลต่อการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวานหรือไม่

$H_0$  : ดัชนีมวลกายที่เปลี่ยนไป ไม่ส่งผลต่อการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน (ไม่สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_0 = 0$





Ha : ระดับไขมันชนิดความหนาแน่นต่ำที่เปลี่ยนไป ส่งผลต่อการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน(สามารถใช้ทำนายตัวแปรตามได้) โดยค่า  $\beta_{11} \neq 0$

3.3.3.4 การแทนค่าที่ไม่สมบูรณ์หรือข้อมูลที่ขาดหายไปด้วยค่าเฉลี่ย ในภาษาไพธอนนั้น สามารถหาค่าเฉลี่ยได้โดยใช้คำสั่ง `np.mean()` ซึ่งจะได้ค่าเฉลี่ยของข้อมูลและสามารถหาค่าที่ไม่สมบูรณ์หรือข้อมูลที่ขาดหายด้วยคำสั่ง `.fillna()` และแทนที่ข้อมูลที่ขาดหายไปด้วยค่าเฉลี่ย

3.3.3.5 การทดสอบระบบ ระบบจะทำการประมวลผลชุดข้อมูลที่กำหนดไว้ด้วยกระบวนการทดสอบระบบ โดยแบ่งการเรียนรู้ของระบบออกเป็น 70% ของข้อมูลทั้งหมด และแบ่งการทดสอบของระบบออกเป็น 30 % ของข้อมูลทั้งหมด

3.3.3.6 การทดสอบระบบ โดยใช้เครื่องมือในการวัดประสิทธิภาพของแบบจำลองดังนี้

- 1) Confusion matrix
- 2) ค่าความถูกต้อง
- 3) ค่าความแม่นยำ
- 4) ค่าความระลึกลับ
- 5) ค่าความถ่วงดุล
- 6) ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ

3.3.3.7 การทดสอบการทำนายผลระบบจะทดสอบการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน โดยมีเกณฑ์การทำนายระยะของโรคไตเรื้อรัง ดังตารางที่ 2.1

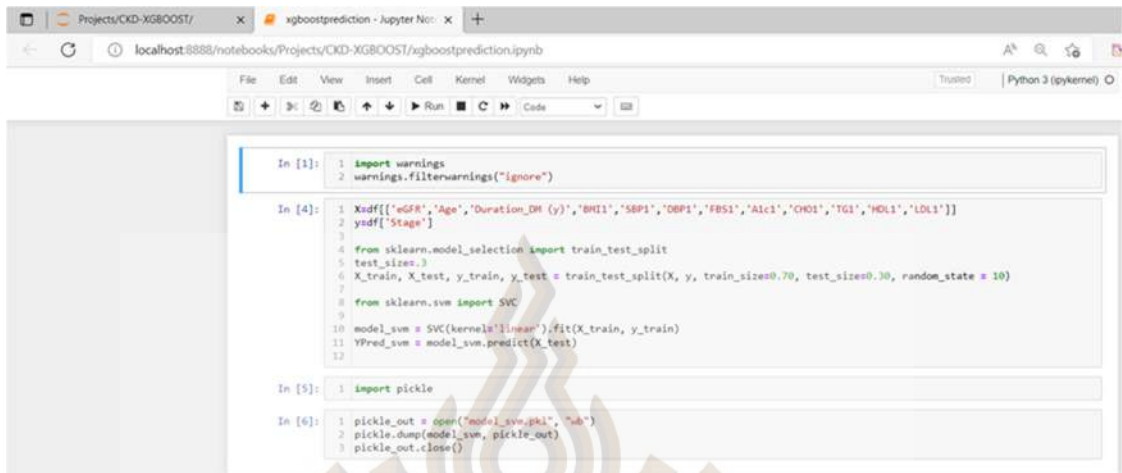
3.3.4 การทดสอบระบบ พัฒนาเว็บแอปพลิเคชันสำหรับตรวจประเมินระยะของโรคในผู้ป่วยโรคไตเรื้อรัง โดยใช้เทคโนโลยีปัญญาประดิษฐ์

จากการวิเคราะห์ระบบมีการพัฒนาเว็บแอปพลิเคชันสำหรับการวิเคราะห์ระยะการเกิดโรคไตเรื้อรังและทำนายระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน

3.3.4.1 สร้างแบบจำลอง โดยการนำข้อมูลของผู้ป่วยโรคเบาหวานที่เป็นโรคไตเรื้อรัง ในโปรแกรม Microsoft Excel กำหนดนามสกุลของไฟล์ Excel เป็น .CSV แล้วนำเข้าสู่โปรแกรม Project Jupyter โดยใช้ภาษาไพธอน ในการเขียนอัลกอริทึมวิธีการซัพพอร์ตเวกเตอร์



แมชชีน นำแบบจำลองที่สร้างขึ้นไปเก็บไปเป็นรูปแบบ .pkl ไฟล์ ด้วยการเรียกใช้ไลบรารี Pickle แสดงดังรูปที่ 3.6



```

In [1]: 1 import warnings
        2 warnings.filterwarnings("ignore")

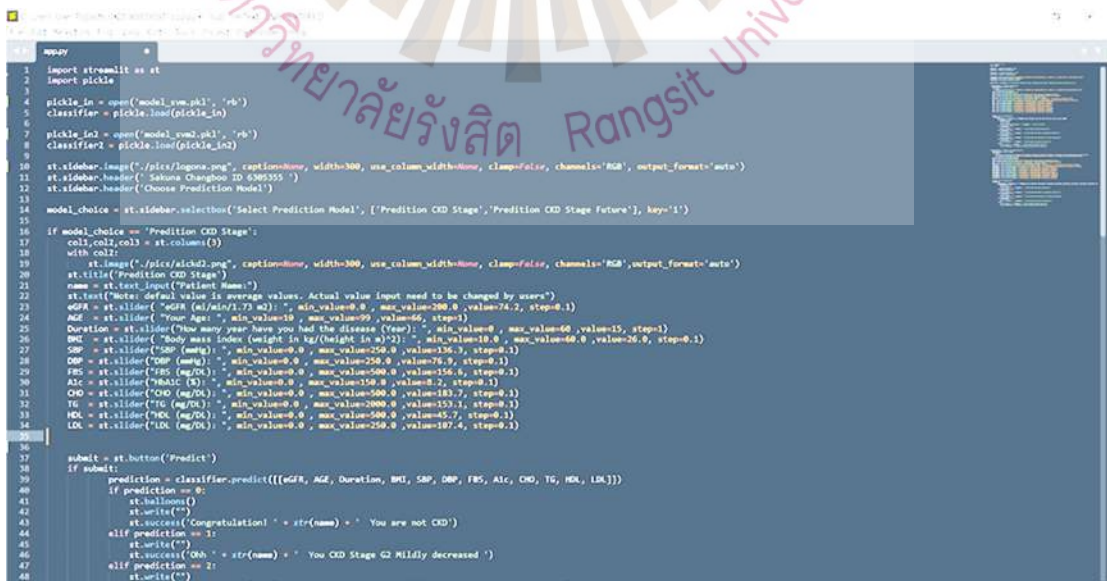
In [4]: 1 Xndf[['eGFR', 'Age', 'Duration_DW (y)', 'BUN1', 'SBP1', 'DBP1', 'FBS1', 'A1c1', 'CHO1', 'TG1', 'HDL1', 'LDL1']]
        2 yndf['Stage']
        3
        4 from sklearn.model_selection import train_test_split
        5 test_size=3
        6 X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.70, test_size=0.30, random_state = 10)
        7
        8 from sklearn.svm import SVC
        9
        10 model_svm = SVC(kernel='linear').fit(X_train, y_train)
        11 yPred_svm = model_svm.predict(X_test)
        12

In [5]: 1 import pickle

In [6]: 1 pickle_out = open("model_svm.pkl", "wb")
        2 pickle.dump(model_svm, pickle_out)
        3 pickle_out.close()
  
```

รูปที่ 3.6 แสดงการเก็บแบบจำลอง

3.3.4.2 สร้างส่วนต่อประสานกับผู้ใช้ (User Interface) เพื่อให้ผู้ใช้งานป้อนข้อมูลคุณลักษณะ ด้วยภาษาไพธอน โดยใช้ไลบรารี Streamlit และเรียกใช้แบบจำลองที่ถูกเก็บไว้ในรูปแบบ .pkl ไฟล์ สำหรับประมวลผลการทำนายจากข้อมูลคุณลักษณะที่ป้อนเข้ามาจากผู้ใช้งาน และแสดงผลพัทธ์การทำนาย แสดงดังรูปที่ 3.7



```

1 import streamlit as st
2 import pickle
3
4 pickle_in = open('model_svm.pkl', 'rb')
5 classifier = pickle.load(pickle_in)
6
7 pickle_in2 = open('model_svm2.pkl', 'rb')
8 classifier2 = pickle.load(pickle_in2)
9
10 st.sidebar.image("./pics/logo.png", caption=None, width=300, use_column_width=None, clamp=False, channels="RGB", output_format="auto")
11 st.sidebar.header("Sattana Changboon ID: 536355")
12 st.sidebar.header("Choose Prediction Model")
13
14 model_choice = st.sidebar.selectbox('Select Prediction Model', ['Prediction CKD Stage', 'Prediction CKD Stage Future', key='1'])
15
16 if model_choice == 'Prediction CKD Stage':
17     col1, col2, col3 = st.columns(3)
18     with col1:
19         st.image("./pics/sichd2.png", caption=None, width=300, use_column_width=None, clamp=False, channels="RGB", output_format="auto")
20     name = st.text_input("Patient Name")
21     st.text("Note: default value is average values. Actual value input need to be changed by users")
22     eGFR = st.slider("eGFR (ml/min/1.73 m2)", min_value=0, max_value=200.0, value=74.2, step=0.1)
23     AGE = st.slider("Your Age", min_value=10, max_value=99, value=66, step=1)
24     Duration = st.slider("How many year have you had the disease (year)", min_value=0, max_value=60, value=15, step=1)
25     BUN = st.slider("Body mass index (weight in kg/(height in m)^2)", min_value=10.0, max_value=60.0, value=26.0, step=0.1)
26     SBP = st.slider("SBP (mmHg)", min_value=0, max_value=210.0, value=135, step=1)
27     DBP = st.slider("DBP (mmHg)", min_value=0, max_value=250.0, value=76.9, step=0.1)
28     FBS = st.slider("FBS (mg/dL)", min_value=0, max_value=200.0, value=155.6, step=0.1)
29     A1c = st.slider("A1c (%)", min_value=0, max_value=15.0, value=7.2, step=0.1)
30     CHO = st.slider("CHO (mg/dL)", min_value=0, max_value=500.0, value=183.7, step=0.1)
31     TG = st.slider("TG (mg/dL)", min_value=0, max_value=2000.0, value=153.1, step=0.1)
32     HDL = st.slider("HDL (mg/dL)", min_value=0, max_value=600.0, value=45.7, step=0.1)
33     LDL = st.slider("LDL (mg/dL)", min_value=0, max_value=250.0, value=107.4, step=0.1)
34
35
36
37 submit = st.button("Predict")
38 if submit:
39     prediction = classifier.predict([[eGFR, AGE, Duration, BUN, SBP, DBP, FBS, A1c, CHO, TG, HDL, LDL]])
40     if prediction == 0:
41         st.success("You are not CKD")
42     elif prediction == 1:
43         st.success("Congratulations! " + str(name) + " You are not CKD")
44     elif prediction == 2:
45         st.success("Oh " + str(name) + " You CKD Stage G2 Mildly decreased")
46     elif prediction == 2:
47         st.success("Oh " + str(name) + " You CKD Stage G2 Mildly decreased")
48     st.write("")
  
```

รูปที่ 3.7 แสดงการสร้างส่วนต่อประสานกับผู้ใช้



### 3.3.5 การประเมินเว็บแอปพลิเคชัน

การประเมินเว็บแอปพลิเคชัน โดยผู้ใช้งาน คือ ผู้เชี่ยวชาญทางการแพทย์และผู้เกี่ยวข้อง จำนวน 6 ท่าน โดยการประเมินมีหัวข้อในการประเมิน 6 ด้าน ได้แก่

- 1) ด้านความสวยงามของเว็บแอปพลิเคชัน
- 2) ด้านการอำนวยความสะดวกภายในเว็บแอปพลิเคชัน
- 3) ด้านประโยชน์จากการทำนายโรคของเว็บแอปพลิเคชัน
- 4) ด้านความรวดเร็วในการตอบสนองการใช้งาน
- 5) ด้านความชัดเจนของข้อมูล
- 6) ด้านความพึงพอใจในการใช้งาน

นำไปสร้างเป็นแบบสอบถามจาก Google Form ดังเอกสารในภาคผนวก ก ให้ผู้ใช้งาน ประเมินและกำหนดการวัดผลความพึงพอใจจะมีเกณฑ์ดังตารางที่ 4.7

ตารางที่ 3.2 เกณฑ์การประเมินผล

ช่วงคะแนน	ระดับความพึงพอใจ
4.50 – 5.00	มากที่สุด
3.50 – 4.49	มาก
2.50 – 3.49	ปานกลาง
1.50 – 2.49	น้อย
1.00 – 1.49	น้อยที่สุด

## บทที่ 4

### ผลการวิจัย

การวิจัยนี้เป็นเป็นการวิเคราะห์ระบบและทำนายความเสี่ยงการเกิดโรคไตเรื้อรังในผู้ป่วยโรคเบาหวาน ซึ่งระบบจะสามารถวิเคราะห์และทำนายได้ว่าผู้ป่วยโรคเบาหวานมีโอกาที่จะเกิดโรคไตเรื้อรังเมื่อไหร่ และระบบสามารถบอกได้ด้วยว่าผู้ป่วยที่เป็นโรคไตเรื้อรังแล้ว ผู้ป่วยจะมีภาวะเสี่ยงที่เป็นไตวายระยะสุดท้ายเร็วขึ้นเท่าไร เพื่อช่วยวิเคราะห์และทำนายผลว่าปัจจัยอะไรที่ทำให้ผู้ป่วยมีค่าอัตราการกรองของไตลดลง โดยรวบรวมและวิเคราะห์ข้อมูลเพื่อหาอัลกอริทึมที่เหมาะสมกับการใช้งานวิจัย

ในการวิจัยครั้งนี้ นอกจากการศึกษาจากทีมผู้วิจัยแล้วผู้วิจัยได้มีการปรึกษาร่วมกับแพทย์ผู้เชี่ยวชาญทางด้านโรคไตเรื้อรังที่ให้การปรึกษาในเรื่องความสัมพันธ์ของข้อมูลผู้ป่วยที่นำมาใช้ในการจัดทำระบบ เพื่อให้ได้ผลการวิจัยที่ถูกต้อง ทางผู้วิจัยจึงได้ขอสรุปการจัดทำโครงการในหัวข้อดังนี้

#### 4.1 ผลการวิจัย

ระบบวิเคราะห์ข้อมูลผู้ป่วยเบาหวานเพื่อทำนายการเกิดโรคไตเรื้อรัง แบ่งผลการจัดทำวิจัยออกเป็น 3 ส่วน ได้แก่ 1) การเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน 2) การวิเคราะห์ระยะการเกิดโรคไตเรื้อรังและทำนายระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวานและ 3) พัฒนาเว็บแอปพลิเคชันสำหรับตรวจประเมินระยะของโรคในผู้ป่วยโรคไตเรื้อรัง โดยใช้เทคโนโลยีปัญญาประดิษฐ์ รวมถึงอภิปรายผลจากการศึกษาข้อดีข้อเสียของเครื่องมือต่าง ๆ ที่นำมาใช้งานวิจัย โดยมีรายละเอียดของผลการวิจัยดังนี้

#### 4.1.1 ผลการเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน

4.1.1.1 ผลการนำข้อมูลสู่เข้าระบบการเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานดังรูปที่ 4.1 ประกอบไปด้วยชุดข้อมูลดังตารางที่ 3.1

จากรูปที่ 4.1 แสดงผลของข้อมูลที่น่าเข้าระบบ ระบบจะแสดงข้อมูลของไฟล์ Excel ที่เป็น .csv เพื่อแสดงถึงชุดข้อมูลที่ต้องใช้ในการประมวลผลการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน

ID	eGFR	Age	uration_DM (	BMI	SBP1	DBP1	FBS1	A1c1	CHO1	TG1	HDL1	LDL1	Stage
10044	98	66	6	24.39	161	85	94	6.6	138	89	34	86	1
10078	98	66	1	28	147	80	148	7.5	175	93	58	106	1
10397	98	57	2	30.49	136	97	287	11.5	246	212	57	147	1
20031	98	57	24	27.21	119	82	155	8.2	138	72	84	48	1
20048	98	71	12	34.72	177	83	106	6.6	128	125	35	68	1
20170	98	73	22	28.92	137	66	78	6.91	200	157	42	127	1
20197	98	69	25	24.88	153	104	152	8.7	177	200	42	95	1
20506	98	75	5	25.82	108	65	107	6.9	154	104	48	85	1
20513	98	73	37	25.93	123	58	150	9	194	165	33	128	1
20601	98	69	20	22.43	146	72	99	6.8	183	114	68	100	1
20637	98	70	20	21.78	145	87	110	7.1	162	107	47	94	1

รูปที่ 4.1 แสดงข้อมูลที่น่าเข้าสู่การวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน

4.1.1.2 ผลการวัดประสิทธิภาพแบบจำลองของอัลกอริทึมการวิเคราะห์การถดถอยโลจิสติก อัลกอริทึมวิธีการค้นหาเพื่อนบ้านใกล้ที่สุด อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน และอัลกอริทึมเอ็กซ์ทรีมกราเดียนบูตตั้ง

เมื่อทำการวิเคราะห์ข้อมูลที่น่าเข้าระบบ พบว่าได้ผลค่า Confusion matrix แสดงดังรูปที่ 4.1

ตารางที่ 4.1 แสดงค่า Confusion Matrix

Confusion matrix	Actual	Prediction	
		eGFR $\geq$ 60 ml/min/1.73 m <sup>2</sup> (Non-CKD)	eGFR < 60 ml/min/1.73 m <sup>2</sup> (CKD)
Logistic Regression	eGFR $\geq$ 60 ml/min/1.73 m <sup>2</sup> (Non-CKD)	302	8
	eGFR < 60 ml/min/1.73 m <sup>2</sup> (CKD)	13	136
K-Nearest Neighbor	eGFR $\geq$ 60 ml/min/1.73 m <sup>2</sup> (Non-CKD)	299	11
	eGFR < 60 ml/min/1.73 m <sup>2</sup> (CKD)	83	66
Support Vector Machine	eGFR $\geq$ 60 ml/min/1.73 m <sup>2</sup> (Non-CKD)	306	4
	eGFR < 60 ml/min/1.73 m <sup>2</sup> (CKD)	7	142
eXtreme Gradient Boosting	eGFR $\geq$ 60 ml/min/1.73 m <sup>2</sup> (Non-CKD)	310	0
	eGFR < 60 ml/min/1.73 m <sup>2</sup> (CKD)	1	148

จากตารางที่ 4.1 แสดงค่า Confusion Matrix เพื่อประเมินผลสัมฤทธิ์การทำงานเปรียบเทียบกับค่าจริง ในรูปแบบตาราง ระหว่างแบบจำลองของแต่ละอัลกอริทึม โดยกำหนดค่าผลลัพธ์ออกเป็น 2 กลุ่ม คือ 0 แสดงถึงผู้ป่วยเบาหวานที่ไม่เป็นโรคไตเรื้อรัง จะมีค่าได้อยู่ในระยะที่ G1 และ G2 คือ มีค่าระดับอัตราการกรองของไตมากกว่าหรือเท่ากับ 60 มิลลิกรัม/นาที/1.73 ตารางเมตร และ 1 แสดงถึงผู้ป่วยเบาหวานที่เป็นโรคไตเรื้อรังอยู่ในระยะที่ G3a, G3b, G4, G5 ซึ่งหมายความว่าผู้ป่วยเบาหวานอยู่ในภาวะโรคไตเรื้อรัง จากนั้นระบบจะทำการวิเคราะห์และตรวจสอบความแม่นยำของระบบ พบว่าประสิทธิภาพการทำนายของแต่ละอัลกอริทึมให้ผลดังนี้

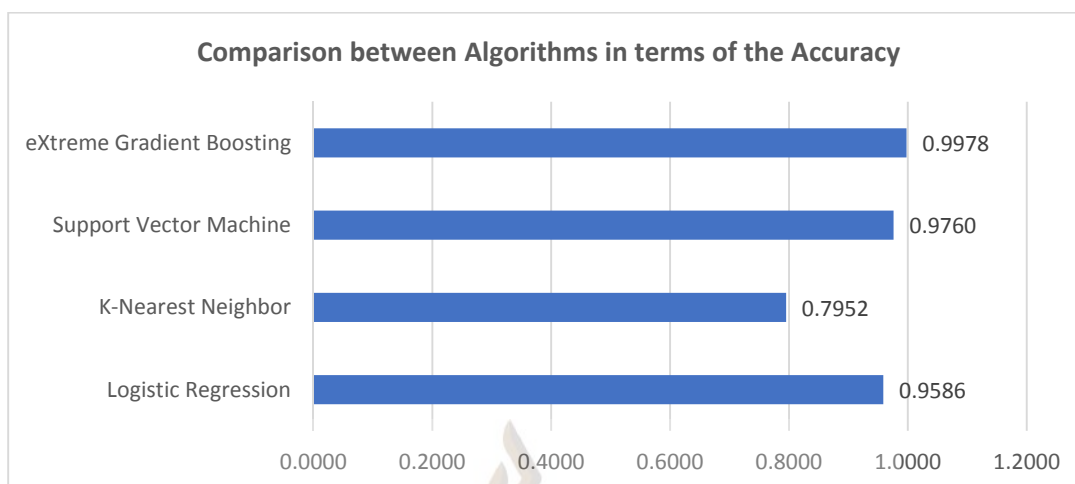
1) อัลกอริทึมการวิเคราะห์การถดถอย โลจิสติก ในกลุ่ม 0 สามารถทำนายได้ถูกต้อง 302 จาก 310 คน และในกลุ่ม 1 ทำนายได้ถูกต้อง 136 จาก 149 คน

2) อัลกอริทึมวิธีการค้นหาเพื่อนบ้านใกล้ที่สุด ในกลุ่ม 0 สามารถทำนายได้ถูกต้อง 299 จาก 310 คน และในกลุ่ม 1 ทำนายได้ถูกต้อง 83 จาก 149 คน

3) อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ในกลุ่ม 0 สามารถทำนายได้ถูกต้อง 306 จาก 310 คน และในกลุ่ม 1 ทำนายได้ถูกต้อง 142 จาก 149 คน

4) อัลกอริทึมเอ็กซ์ทรีมกราดิเอนท์บูสติง ในกลุ่ม 0 สามารถทำนายได้ถูกต้อง 310 จาก 310 คน และในกลุ่ม 1 ทำนายได้ถูกต้อง 148 จาก 149 คน

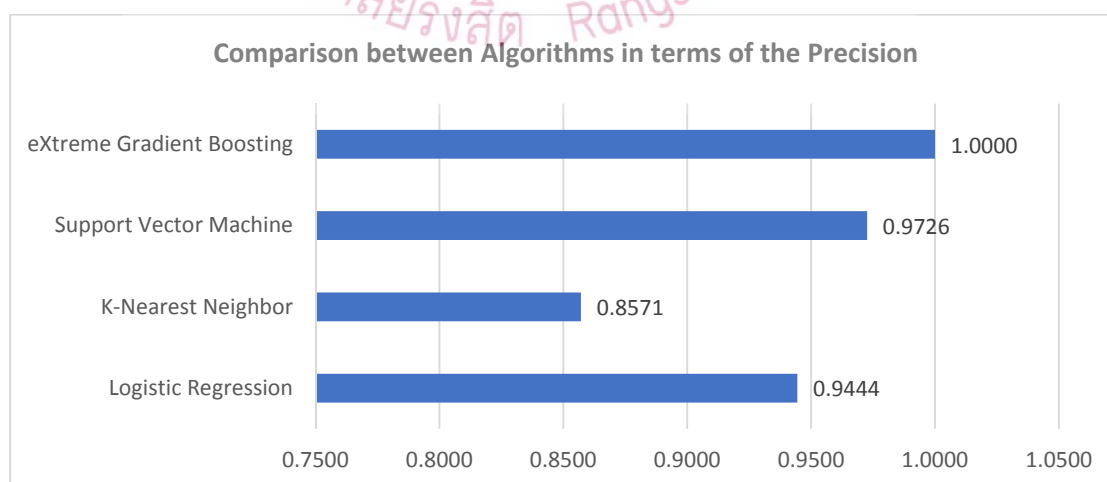
เมื่อเปรียบเทียบแต่ละอัลกอริทึมโดยพิจารณาจากค่าความถูกต้อง สามารถแสดงได้ดังรูปที่ 4.2



รูปที่ 4.2 แสดงกราฟเปรียบเทียบประสิทธิภาพด้านความถูกต้องของแต่ละอัลกอริทึม

จากรูปที่ 4.2 แสดงการเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานในด้านความถูกต้อง พบว่า อัลกอริทึมเอ็กซ์ทรีมกราเดียนบูสต์ ให้ค่าความถูกต้อง สูงที่สุดคือ 0.9978 รองลงมาเป็นอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ให้ค่า ความถูกต้อง 0.9760 อัลกอริทึมการวิเคราะห์การถดถอยโลจิสติก ให้ค่าความถูกต้อง 0.9586 และ สุดท้ายอัลกอริทึมการค้นหาเพื่อนบ้านใกล้ที่สุด ให้ค่าความถูกต้องน้อยที่สุด 0.7952 ตามลำดับ

เมื่อเปรียบเทียบแต่ละอัลกอริทึมโดยพิจารณาจากด้านความแม่นยำ สามารถแสดงได้ดังรูปที่ 4.3

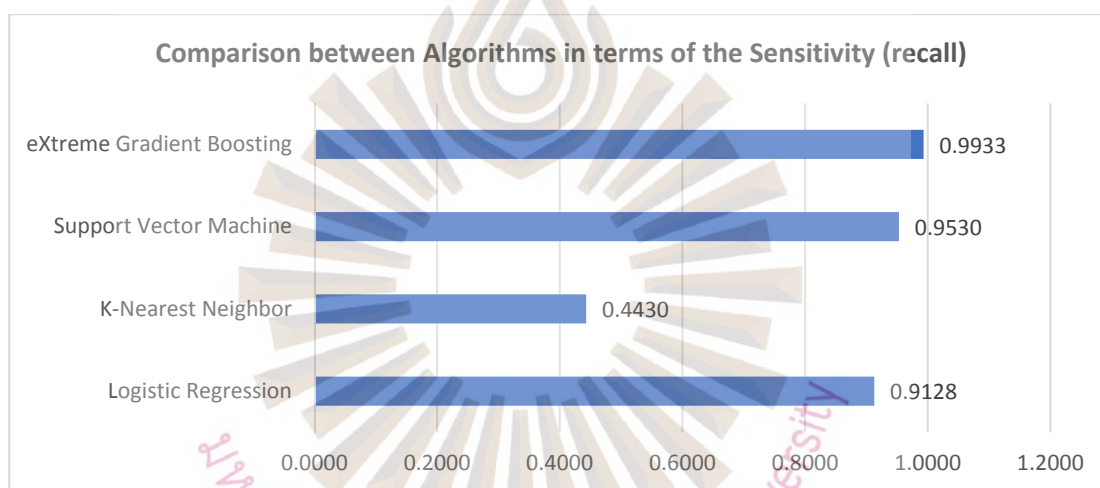


รูปที่ 4.3 แสดงกราฟเปรียบเทียบประสิทธิภาพด้านความแม่นยำของแต่ละอัลกอริทึม



จากรูปที่ 4.1 แสดงการเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานในด้านความแม่นยำ กล่าวคือเป็นอัตราส่วนของการค้นพบข้อมูลที่ถูกต้องจากจำนวนข้อมูลทั้งหมดที่ทำการค้นคืนมาได้ พบว่า อัลกอริทึมเอ็กซ์ทรีมกราเดียนบูตติ้ง ให้ค่าความแม่นยำ สูงที่สุดคือ 1.0000 รองลงมาเป็น อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความแม่นยำ 0.9726 อัลกอริทึมการวิเคราะห์การถดถอยโลจิสติก ให้ค่าความแม่นยำ 0.9444 และสุดท้าย อัลกอริทึมการค้นหาเพื่อนบ้านใกล้ที่สุด ให้ค่าความแม่นยำ น้อยที่สุด 0.8571 ตามลำดับ

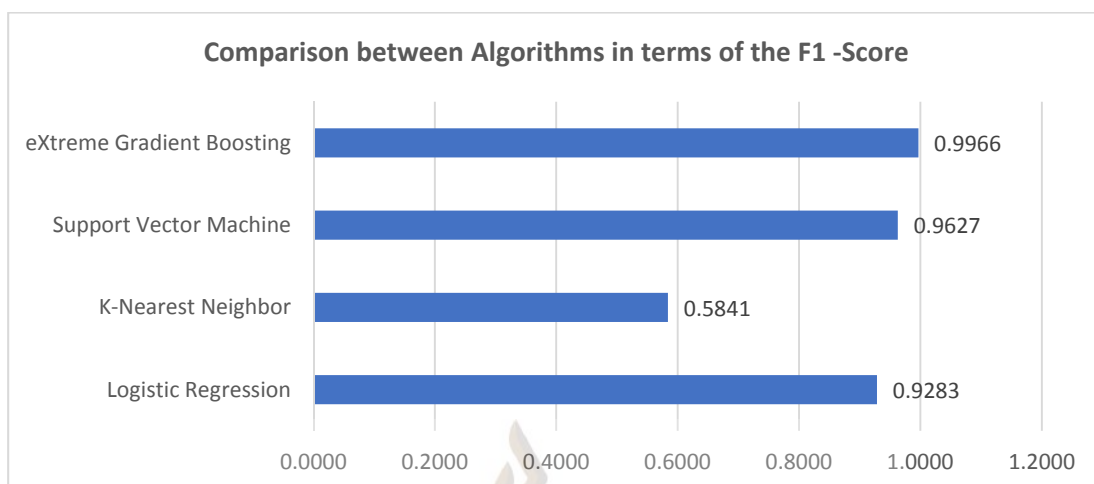
เมื่อเปรียบเทียบแต่ละอัลกอริทึม โดยพิจารณาจากค่าความระลึก สามารถแสดงได้ ดังรูปที่ 4.4



รูปที่ 4.4 แสดงกราฟเปรียบเทียบประสิทธิภาพด้านค่าเรียกกลับของแต่ละอัลกอริทึม

จากรูปที่ 4.4 แสดงการเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานในด้านค่าความระลึก กล่าวคือเป็นอัตราส่วนของการค้นพบข้อมูลที่ถูกต้องจากจำนวน ข้อมูลที่ถูกต้องทั้งหมด พบว่า อัลกอริทึมเอ็กซ์ทรีมกราเดียนบูตติ้ง ให้ค่าความระลึก สูงที่สุดคือ 0.9933 รองลงมาเป็น อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความระลึก 0.9530 อัลกอริทึมการวิเคราะห์การถดถอยโลจิสติก ให้ค่าความระลึก 0.9444 และสุดท้าย อัลกอริทึมการค้นหาเพื่อนบ้านใกล้ที่สุด ให้ค่าความระลึกน้อยที่สุด 0.8571 ตามลำดับ

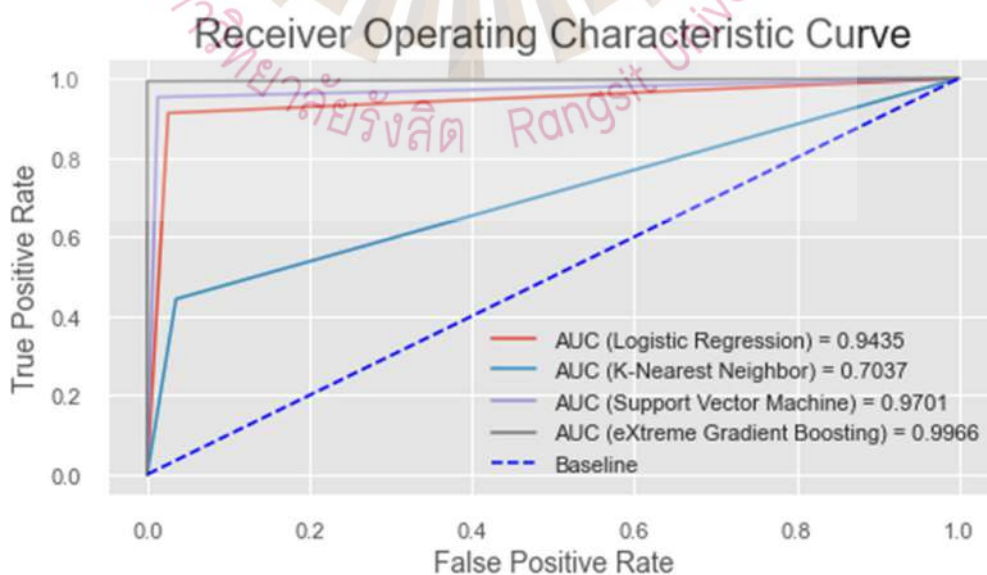
เมื่อเปรียบเทียบแต่ละอัลกอริทึม โดยพิจารณาจากค่าความถ่วงดุล สามารถแสดงได้ ดังรูปที่ 4.5



รูปที่ 4.5 แสดงกราฟเปรียบเทียบประสิทธิภาพด้านค่าความถ่วงดุลของแต่ละอัลกอริทึม

จากรูปที่ 4.4 แสดงการเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานในด้านค่าความถ่วงดุล พบว่าอัลกอริทึมเอ็กซ์ตรีมกราเดียนบูสติ่ง ให้ค่าความถ่วงดุล สูงที่สุดคือ 0.9966 รองลงมาเป็น อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความถ่วงดุล 0.9627 อัลกอริทึมการวิเคราะห์การถดถอยโลจิสติก ให้ค่าความถ่วงดุล 0.9283 และสุดท้าย อัลกอริทึมการค้นหาเพื่อนบ้านใกล้ที่สุด ให้ค่าความถ่วงดุล 0.5841 ตามลำดับ

เมื่อเปรียบเทียบแต่ละอัลกอริทึม โดยพิจารณาจากค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ สามารถแสดงได้ดังรูปที่ 4.5



รูปที่ 4.6 แสดงกราฟเปรียบเทียบประสิทธิภาพด้วยค่าตัววัดประสิทธิภาพ โดยใช้พื้นที่ใต้กราฟ ROC Curve

จากรูปที่ 4.5 แสดงการเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน ด้วยค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ กล่าวคือกราฟแสดงความสัมพันธ์ ของ แกน Y และ แกน X โดยให้การทำนายถูกต้องไปทางแกนนตั้ง (Y) และถ้าทำนายผิดไปแนวแกนนอน (X) และถ้าค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ มีค่าเข้าใกล้ 1 จะแสดงว่าแบบจำลองนั้นมีประสิทธิภาพที่ดีกว่า ผลการทดลองพบว่าอัลกอริทึมเอ็กซ์ทรีมกราเดียนบูตดิ้ง ให้ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ สูงที่สุดคือ 0.9966 อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าตัววัดประสิทธิภาพ โดยใช้พื้นที่ใต้กราฟ 0.9701 อัลกอริทึมการวิเคราะห์การถดถอยโลจิสติก ให้ค่าตัววัดประสิทธิภาพ โดยใช้พื้นที่ใต้กราฟ 0.9435 และสุดท้าย อัลกอริทึมการค้นหาเพื่อนบ้านใกล้ที่สุด ให้ค่าตัววัดประสิทธิภาพ โดยใช้พื้นที่ใต้กราฟ 0.7037 ตามลำดับ

สรุปถึงแม้ว่าอัลกอริทึมเอ็กซ์ทรีมกราเดียนบูตดิ้ง ให้ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ สูงที่สุดคือ 0.9966 หรือ 1.00 เป็นค่าที่ทำให้เกิด Overfitting คือ กรณีการที่แบบจำลองมีความแม่นยำสูงมากในตอนทดลองแต่พอเอาไปใช้งานจริงกลับได้ผลลัพธ์ที่ไม่ดีของระบบได้ ดังนั้นจึงเลือกใช้ อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ที่ให้ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ 0.9701 ในการวิเคราะห์ระยะการเกิดโรคไตเรื้อรังและทำนายระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวานต่อไป

#### 4.1.2 ผลการสร้างแบบจำลองที่ใช้ในการวิเคราะห์ระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน

4.1.2.1 ผลการจัดการข้อมูลที่น่าเข้าระบบการวิเคราะห์ระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานดังรูปที่ 4.2 ประกอบไปด้วยชุดข้อมูลดังตารางที่ 3.1

ID	eGFR	Age	uration_DM (	BMI1	SBP1	DBP1	FBS1	A1c1	CHO1	TG1	HDL1	LDL1	Stage
10044	90	66	6	24.39	161	85	94	6.6	138	89	34	86	1
10078	90	66	1	28	147	88	148	7.5	175	93	58	166	1
10397	90	57	2	30.49	136	97	287	11.5	246	212	57	147	1
20031	90	57	24	22.21	119	82	155	8.2	138	72	84	48	1
20048	90	71	12	34.22	127	83	106	6.6	128	125	35	68	1
20170	90	73	22	28.82	137	66	78	6.91	200	157	42	127	1
20197	90	69	25	24.88	153	104	152	8.7	177	200	42	95	1
20506	90	75	5	25.82	108	65	107	6.9	154	184	48	85	1
20513	90	73	37	25.03	123	58	150	9	194	165	33	128	1
20601	90	69	28	22.43	146	72	99	6.8	183	114	68	100	1
20637	90	70	20	21.76	145	87	110	7.1	162	107	47	94	1

รูปที่ 4.7 แสดงข้อมูลที่น่าเข้าสู่การวิเคราะห์ระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน

จากรูปที่ 4.7 แสดงผลของข้อมูลที่น่าเข้าระบบ ระบบจะแสดงข้อมูลของไฟล์ Excel ที่กำหนดนามสกุลของไฟล์เป็น .csv เพื่อแสดงถึงชุดข้อมูลที่ต้องใช้ในการประมวลผลการวิเคราะห์ระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน

#### 4.1.2.2 ผลการทำความสะอาดข้อมูล

แสดงการแทนค่าที่ไม่สมบูรณ์หรือข้อมูลที่ขาดหายด้วยค่าเฉลี่ยของข้อมูลจะได้ผลลัพธ์เป็นชุดข้อมูลใหม่ดังแสดงในรูปที่ 4.8

```
RangeIndex: 1528 entries, 0 to 1527
Data columns (total 14 columns):
#   Column                               Non-Null Count  Dtype
---  ---                               -
0   ID                                     1528 non-null   int64
1   eGFR                                  1528 non-null   float64
2   Age                                   1528 non-null   int64
3   Duration_DM (y)                       1528 non-null   int64
4   BMI1                                  1527 non-null   float64
5   SBP1                                  1528 non-null   int64
6   DBP1                                  1528 non-null   int64
7   FBS1                                  1528 non-null   int64
8   A1c1                                  1528 non-null   float64
9   CHO1                                  1528 non-null   int64
10  TG1                                   1528 non-null   int64
11  HDL1                                  1528 non-null   int64
12  LDL1                                  1528 non-null   int64
13  Stage                                 1528 non-null   int64
dtypes: float64(3), int64(11)
memory usage: 167.2 KB
```

(ก) แสดงชนิดของข้อมูลแต่ละข้อมูล

```
dtypes: float64(3), int64(11)
memory usage: 167.2 KB
None
ID                                     0
eGFR                                  0
Age                                   0
Duration_DM (y)                       0
BMI1                                  1
SBP1                                  0
DBP1                                  0
FBS1                                  0
A1c1                                  0
CHO1                                  0
TG1                                   0
HDL1                                  0
LDL1                                  0
Stage                                 0
dtype: int64
```

(ข) แสดงผลค่า Nan ในระบบ

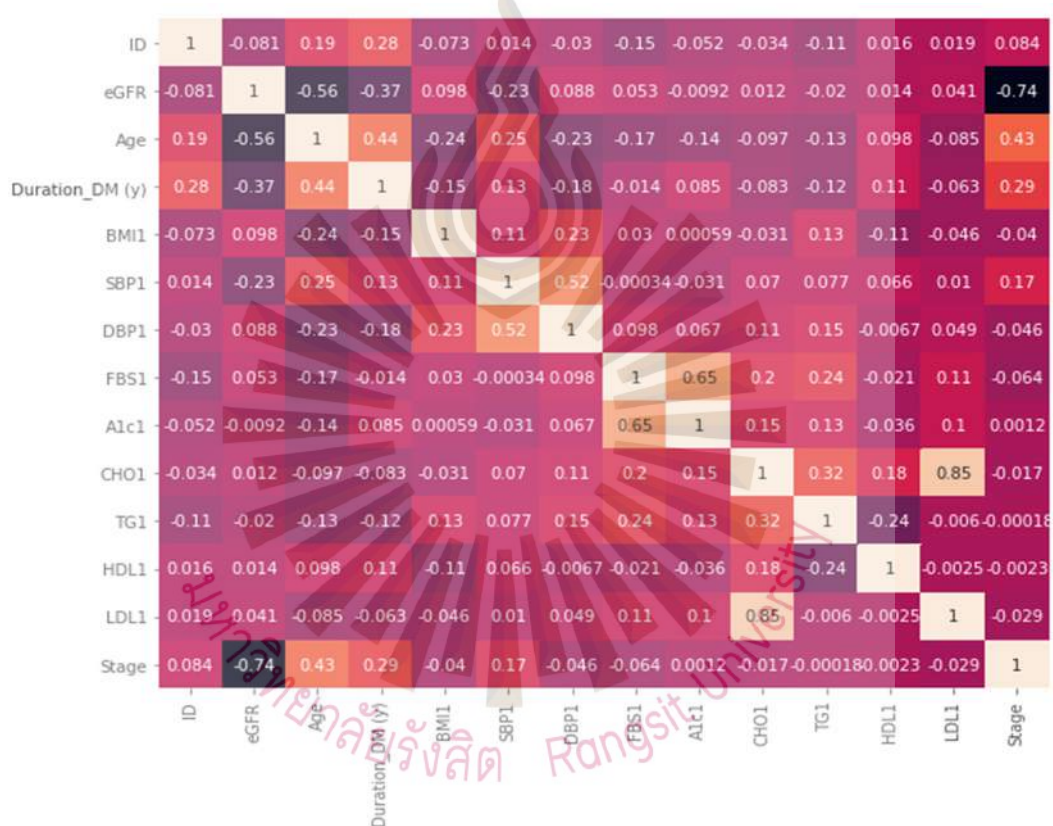
รูปที่ 4.8 แสดงลักษณะและความมีอยู่ของข้อมูลนำเข้าระบบ ชุดที่ 1



จากรูปที่ 4.8 (ก) บอกรูปของข้อมูลแต่ละข้อมูล เช่น int64 คือตัวเลข float64 คือตัวเลขที่เป็นจุดทศนิยม และ Object ที่เป็นข้อมูลวันที่ (ข) แสดงค่า nan เป็นค่าว่าง ไม่ได้นำมาใช้งานระบบจึงทำการล้างออกแล้วแทนด้วยค่าเฉลี่ยของข้อมูล

#### 4.1.2.3 ผลการหาค่าความสัมพันธ์ของข้อมูลสุขภาพ

เมื่อทดสอบความสัมพันธ์ของข้อมูลแต่ละตัวโดยพิจารณาจากข้อมูลผลลัพธ์คือ Stage สามารถแสดงได้ดังรูปที่ 4.9



รูปที่ 4.9 แสดงความสัมพันธ์ของข้อมูลแต่ละตัว

จากรูปที่ 4.9 แสดงถึงค่าสัมประสิทธิ์สหสัมพันธ์ เป็นตัวบ่งชี้ถึงความสัมพันธ์นี้ ซึ่งค่าสัมประสิทธิ์สหสัมพันธ์นี้จะมีค่าอยู่ระหว่าง -1.0 ถึง + 1.0 ซึ่งหากมีค่าใกล้ -1.0 นั้นหมายความว่าตัวแปรทั้งสองตัวมีความสัมพันธ์กันในเชิงตรงกันข้ามอย่างมาก หากมีค่าใกล้ +1.0 นั้นหมายความว่าตัวแปรทั้งสองมีความสัมพันธ์กันโดยตรงอย่างมากและหากมีค่าเป็น 0 นั้นหมายความว่าตัวแปรทั้งสองตัวไม่มีความสัมพันธ์ต่อกัน รวมถึงเลือกใช้ข้อมูลที่มีความสัมพันธ์ต่อการวิเคราะห์ระยะการ



เกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานตามทฤษฎีทางการแพทย์ เมื่อกำหนดข้อมูลผลลัพธ์ คือ Stage พบว่าข้อมูลคุณลักษณะ ที่สำคัญ 3 อันดับแรกของการทำนายผลลัพธ์ ได้แก่ อัตราการกรองของไต ที่ค่า 0.74 อายุ ที่ค่า 0.43 และระยะเวลาที่ผู้ป่วยเป็นโรคมมาแล้วกี่ปี ที่ค่า 0.29 ตามลำดับ

#### 4.1.2.4 ผลการหาค่าสำคัญทางสถิติของข้อมูลสุขภาพ

เมื่อทำการทดสอบสมมติฐานทางสถิติของข้อมูลสุขภาพ พบว่าได้ผลค่านัยสำคัญ แสดงดังตารางที่ 4.2

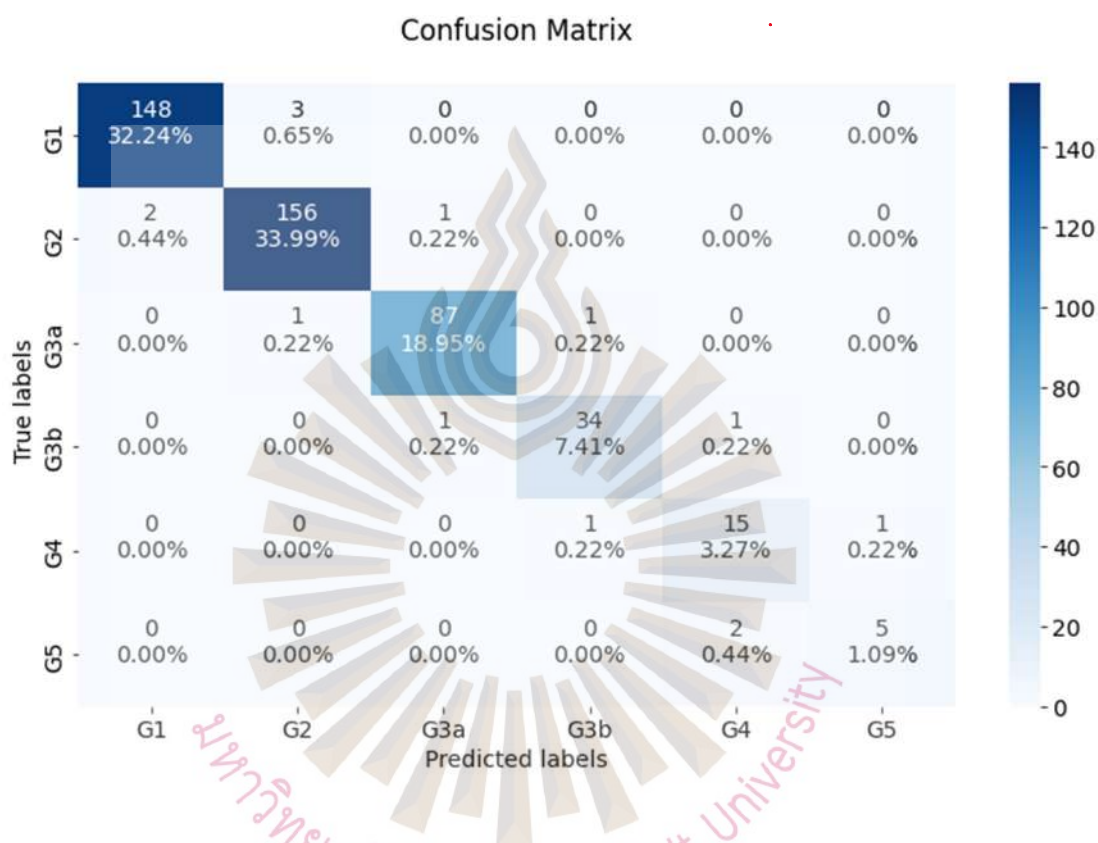
ตารางที่ 4.2 แสดงนัยสำคัญทางสถิติของข้อมูลสุขภาพ

ลำดับ	ข้อมูล	P-Value
1	Age (Year)	1.049
2	Duration_DM (Year)	0.378
3	BMI1 (kg/m <sup>2</sup> )	0.124
4	SBP1 (mmHg)	0.037
5	HDL1 (mg/DL)	0.351
6	DBP1 (mmHg)	0.138
7	FBS1 (mg/DL)	0.045
8	HbA1C (%)	0.298
9	CHO1 (mg/DL)	0.040
10	TG1 (mg/DL)	0.016
11	LDL1 (mg/DL)	0.025

จากตารางที่ 4.2 จากค่า P-Value มีระดับนัยสำคัญทางสถิติน้อยกว่า 0.05 หรือ 0.025 และค่า P-Value บางค่ามีระดับนัยสำคัญทางสถิติมากกว่า 0.05 แสดงให้เห็นว่า ผลที่ได้จากการศึกษาไม่เพียงพอที่จะสามารถนำมาใช้ในการทดสอบระบบ เนื่องจากสมมติฐานแสดงถึงความสัมพันธ์ของข้อมูลตัวแปรตาม (Y) ไม่มีความสัมพันธ์กับตัวแปรอิสระ (X) เมื่อกำหนดตัวแปรอิสระ และกำหนดตัวแปรตามแล้ว แต่ตัวแปรอิสระ (X) บางตัวสามารถนำมาใช้ในการทดสอบค่าระดับนัยสำคัญของตัวแปรตามได้ จึงใช้ข้อมูลในการเรียนรู้และทดสอบระบบ นอกจากนี้มีตัวแปรอิสระที่ไม่มีความสัมพันธ์กับข้อมูลแต่ต้องนำมาใช้ในการทดสอบระบบเนื่องจากมีความสัมพันธ์กันในทางการแพทย์

4.1.2.5 ผลการวัดประสิทธิภาพแบบจำลองของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ในการวิเคราะห์ระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน

เมื่อทำการวิเคราะห์ข้อมูลที่น่าจะเข้าระบบ พบว่าได้ผลค่า Confusion Matrix แสดงดังรูปที่ 4.10

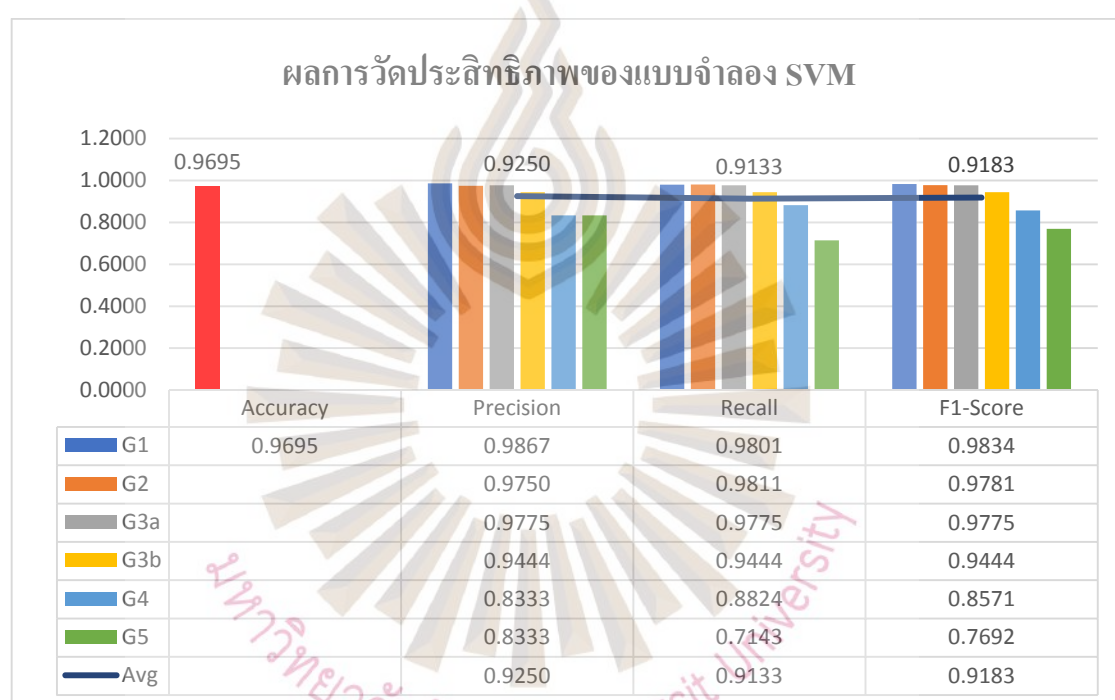


รูปที่ 4.10 แสดงค่า Confusion Matrix ของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน

จากรูปที่ 4.10 แสดงค่า Confusion Matrix เป็นตารางที่ใช้ประเมินผลลัพธ์การทำนาย เปรียบเทียบกับค่าจริง ของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน โดยแบ่ง ข้อมูลผลลัพธ์ ออกเป็น 6 กลุ่ม คือแสดงถึงผู้ป่วยเบาหวานที่เป็นโรคไตเรื้อรังในซึ่งระยะ G1 มีค่าระดับอัตราการกรองของไตมากกว่าหรือเท่ากับ 90 มิลลิลิตร/นาที/1.73 ตารางเมตร ระยะ G2 มีค่าระดับอัตราการกรองของไตระหว่าง 60-89 มิลลิลิตร/นาที/1.73 ตารางเมตร ระยะ G3a มีค่าระดับอัตราการกรองของไตระหว่าง 45-59 มิลลิลิตร/นาที/1.73 ตารางเมตร ระยะ G3b มีค่าระดับอัตราการกรองของไตระหว่าง 30-44 มิลลิลิตร/นาที/1.73 ตารางเมตร ระยะ G4 มีค่าระดับอัตราการกรองของไตระหว่าง 15-29 มิลลิลิตร/นาที/1.73 ตารางเมตร ระยะ G5 มีค่าระดับอัตราการกรองของไตน้อยกว่า 15 มิลลิลิตร/นาที/1.73 ตารางเมตร จากนั้นระบบจะทำการวิเคราะห์และตรวจสอบความแม่นยำของ

ระบบ พบว่า ระยะ G1 ทำนายได้ถูกต้อง 148 จาก 151 คน หรือเท่ากับ 32.24 % ระยะ G2 ทำนายได้ถูกต้อง 156 จาก 159 คน หรือเท่ากับ 33.99 % ระยะ G3a ทำนายได้ถูกต้อง 87 จาก 89 คน หรือเท่ากับ 18.95% ระยะ G3b ทำนายได้ถูกต้อง 34 จาก 36 คน หรือเท่ากับ 7.41 % ระยะ G4 ทำนายได้ถูกต้อง 15 จาก 17 คน หรือเท่ากับ 3.27% และระยะ G5 ทำนายได้ถูกต้อง 5 จาก 7 คน หรือเท่ากับ 1.09 %

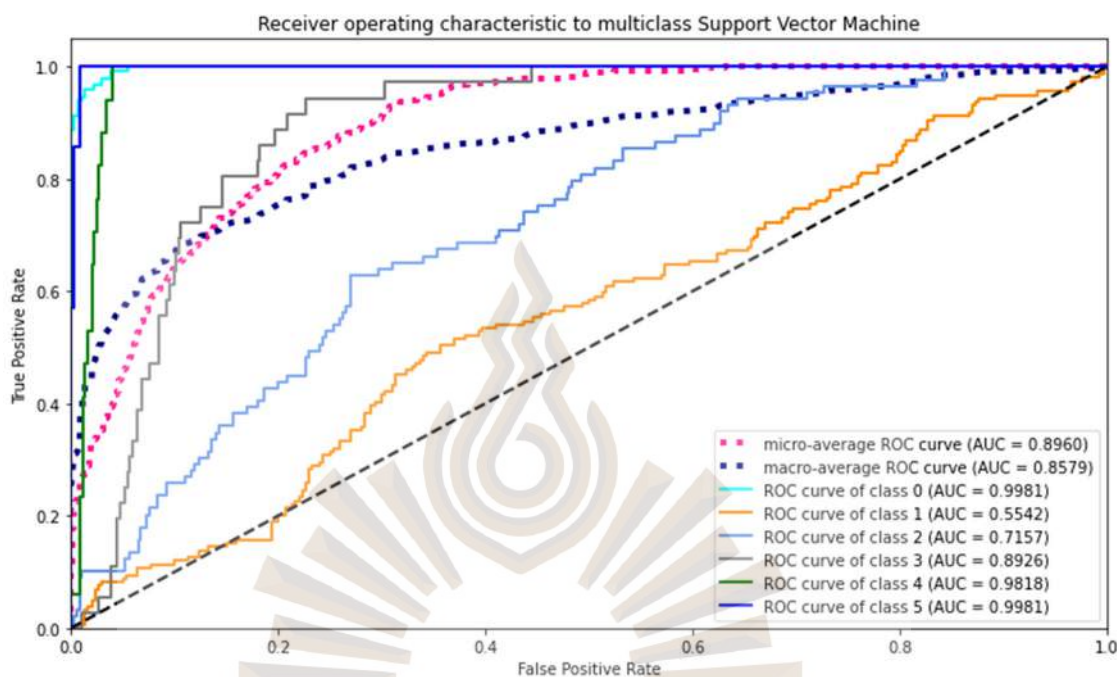
เมื่อทำการวิเคราะห์ข้อมูลที่น่าเข้าระบบพบว่า ได้ผลค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก ค่าความถ่วงดุล แสดงดังรูปที่ 4.11



รูปที่ 4.11 กราฟแสดงผลการวัดประสิทธิภาพของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน

จากรูปที่ 4.11 แสดงผลการวัดประสิทธิภาพของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน มีค่าความถูกต้องทั้งหมด อยู่ที่ 0.9695 ค่าความแม่นยำ ระยะ G1 ให้ค่า 0.9867 ระยะ G2 ให้ค่า 0.9750 ระยะ G3a ให้ค่า 0.9775 ระยะ G3b ให้ค่า 0.9444 ระยะ G4 ให้ค่า 0.8333 ระยะ G5 ให้ค่า 0.8333 ค่าความระลึก ระยะ G1 ให้ค่า 0.9801 ระยะ G2 ให้ค่า 0.9811 ระยะ G3a ให้ค่า 0.9775 ระยะ G3b ให้ค่า 0.9444 ระยะ G4 ให้ค่า 0.8824 ระยะ G5 ให้ค่า 0.7143 ค่าความถ่วงดุล ระยะ G1 ให้ค่า 0.9834 ระยะ G2 ให้ค่า 0.9781 ระยะ G3a ให้ค่า 0.9775 ระยะ G3b ให้ค่า 0.9444 ระยะ G4 ให้ค่า 0.8571 และระยะ G5 ให้ค่า 0.7692

เมื่อเปรียบเทียบแต่ละอัลกอริทึม โดยพิจารณาจากค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ ROC Curve สามารถแสดงได้ดังรูปที่ 4.12



รูปที่ 4.12 แสดงพื้นที่ใต้กราฟ ROC Curve ของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน โดยแบ่งผลลัพธ์เป็น 6 กลุ่ม

จากรูปที่ 4.12 แสดงผลการวัดประสิทธิภาพของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ด้วยค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ โดยกราฟแสดงความสัมพันธ์ การทำนายถูกต้องไปทางแกนตั้ง (Y) และถ้าทำนายผิดไปแนวแกนนอน (X) และถ้าค่า AUCs มีค่าเข้าใกล้ 1 จะแสดงว่า ข้อมูลผลลัพธ์ในกลุ่มนั้นมีประสิทธิภาพที่ดี โดยระยะ G1 ให้ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ 0.9981 ระยะ G2 ให้ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ 0.5542 ระยะ G3a ให้ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ 0.7157 ระยะ G3b ให้ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ 0.8926 ระยะ G4 ให้ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ 0.9818 ระยะ G5 ให้ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ 0.9981



### 4.1.3 ผลการสร้างแบบจำลองที่ใช้ในการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน

#### 4.1.3.1 ผลการนำข้อมูลที่น่าเข้าสู่การทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวานดังรูปที่ 4.3 ประกอบไปด้วยชุดข้อมูลดังตารางที่ 3.2

Vstdate_first	Vstime_first	eGFR1	/stime_chanc	eGFR2	luration(day	Age	DurationY	BMLchange	SBP_shange	DBP_change	FBS_change	A1c_change	CHO_change	TG_change	HDL_ch
9/5/2006	16/18/2012	42.7	29/3/2016	53	1268	76	19	21.85	161	98	167	6.9	182	235	28
9/7/2007	11/12/2012	24.2	1/1/2016	11	1116	81	39	27.86	168	89	122	7.7	151	179	36
25/12/2009	18/18/2014	55	26/3/2015	68	167	79	18	26.8924	137	102	83	8.2846	177.816	163.597	47.5
27/7/2006	18/7/2012	51.7	12/2/2013	44.3	217	64	11	26.8924	98	76	172	8.3	179	128	43
26/12/2013	26/12/2013	51	38/1/2014	79	35	74	19	26.8924	204	114	248	11.6	157	178	25
23/1/2013	6/6/2014	58	17/9/2015	53	468	74	17	26.8924	116	68	148	8.4	185	212	46
1/2/2005	28/6/2011	15.5	11/12/2012	28	548	58	26	26.8924	117	84	119	18.7	158	142	35
2/12/2013	2/12/2013	48	8/9/2015	38	645	62	7	26.8924	133	71	188	9.1	197	158	57
2/11/2009	23/4/2014	88	27/18/2015	67	552	62	7	26.8924	148.566	131	178	9.3	132	143	32
24/18/2006	25/12/2012	32.8	17/9/2013	42	266	61	26	26.8924	131	63	167	8.5	326	227	48
2/3/2012	11/18/2012	43.2	17/18/2013	58	371	68	9	26.8924	12	62	123	5.9	133	91	49
16/9/2013	16/9/2013	38	31/7/2014	29	318	68	17	26.8924	141	78	138	7.7	211	186	61
9/2/2005	26/7/2011	29.5	22/11/2012	16.6	485	74	15	26.8924	131	68	252	7.6	181	185	48
15/3/2005	9/18/2012	31.5	5/11/2013	53	392	53	5	26.8924	151	98	151	98	194	145	38
38/12/2005	28/12/2012	32.9	26/18/2015	45	1832	87	58	26.8924	122	51	178	7.8	187	163	62
18/2/2018	8/11/2012	32.9	8/8/2013	26	273	86	12	26.8924	156	74	129	7.5	118	113	58
11/1/2005	18/12/2012	48.2	17/12/2013	55	364	75	23	26.8924	148.566	75.3883	151	7.5	172	182	46
19/4/2005	11/18/2012	31.9	1/8/2013	44	294	49	13	26.8924	141	85	132	8.4	256	291	42
8/3/2005	8/11/2012	48.4	18/11/2013	43	375	73	22	26.8924	146	78	155.823	8.4	132	77	34
21/12/2009	27/11/2012	48.8	17/12/2013	53	385	78	9	26.8924	153	76	11	8.5	176	132	31
26/12/2005	27/11/2012	42.4	17/12/2013	63	385	75	11	26.8924	157	79	187	8.2846	177.816	163.597	47.5

รูปที่ 4.13 แสดงข้อมูลที่น่าเข้าสู่การทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน โดยใช้ข้อมูลของผู้ป่วยโรคเบาหวาน

จากรูปที่ 4.13 แสดงผลของข้อมูลที่น่าเข้าระบบ ระบบจะแสดงข้อมูลของไฟล์ Excel ที่กำหนดนามสกุลของไฟล์เป็น .csv เพื่อแสดงถึงชุดข้อมูลที่ควรใช้ในการประมวลผลการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน โดยใช้ข้อมูลของผู้ป่วยโรคเบาหวาน

#### 4.1.3.2 ผลการทำความสะอาดข้อมูล

แสดงการแทนค่าที่ไม่สมบูรณ์หรือข้อมูลที่ขาดหายด้วยค่าเฉลี่ยของข้อมูลจะได้ผลลัพธ์เป็นชุดข้อมูลใหม่ดังแสดงในรูปที่ 4.14



```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 440 entries, 0 to 439
Data columns (total 20 columns):
#   Column              Non-Null Count  Dtype
---  -
0   ID                   440 non-null    int64
1   Vstdate_first        154 non-null    object
2   Vstime_first         440 non-null    object
3   eGFR1                440 non-null    float64
4   Vstime_chang         440 non-null    object
5   eGFR2                440 non-null    float64
6   stated_change        440 non-null    int64
7   durasion(day)        440 non-null    int64
8   Age                  440 non-null    int64
9   DurationY            440 non-null    int64
10  BMI_change            231 non-null    float64
11  SBP_shange           373 non-null    float64
12  DBP_change           421 non-null    float64
13  FBS_change           429 non-null    float64
14  A1c_change           413 non-null    float64
15  CHO_change           426 non-null    float64
16  TG_channge          425 non-null    float64
17  HDL_change           424 non-null    float64
18  LDL_change           294 non-null    float64
19  Stage                433 non-null    float64
dtypes: float64(12), int64(5), object(3)
memory usage: 68.9+ KB

```

```

None
ID                   0
Vstdate_first       286
Vstime_first         0
eGFR1                0
Vstime_chang         0
eGFR2                0
stated_change        0
durasion(day)        0
Age                  0
DurationY            0
BMI_change           209
SBP_shange           67
DBP_change           19
FBS_change           11
A1c_change           27
CHO_change           14
TG_channge           15
HDL_change           16
LDL_change           146
Stage                7
dtype: int64

```

(ก) แสดงชนิดของข้อมูลแต่ละข้อมูล

(ข) แสดงค่า Nan ในระบบ

รูปที่ 4.14 แสดงลักษณะและความมีอยู่ของข้อมูลนำเข้าระบบ ชุดที่ 2

จากรูปที่ 4.14 (ก) บอกชนิดของข้อมูลแต่ละข้อมูล เช่น int64 คือตัวเลข float64 คือตัวเลขที่เป็นจุดทศนิยมและ Object ที่เป็นข้อมูลวันที่ (ข) แสดงค่า Nan เป็นค่าว่าง ไม่ได้นำมาใช้งานระบบจึงทำการล้างออกแล้วแทนด้วยค่าเฉลี่ยของข้อมูล

#### 4.1.3.3 ผลการหาค่าความสัมพันธ์ของข้อมูลสุขภาพ

เมื่อทดสอบความสัมพันธ์ของข้อมูลแต่ละตัวโดยพิจารณาจากข้อมูลผลลัพธ์คือ eGFR2สามารถแสดงได้ดังรูปที่ 4.15



รูปที่ 4.15 แสดงความสัมพันธ์ของข้อมูล

จากรูปที่ 4.15 แสดงถึงค่าสัมประสิทธิ์สหสัมพันธ์ เป็นตัวบ่งชี้ถึงความสัมพันธ์นี้ ซึ่งค่าสัมประสิทธิ์สหสัมพันธ์นี้มีค่าอยู่ระหว่าง -1.0 ถึง +1.0 ซึ่งหากมีค่าใกล้ -1.0 นั้นหมายความว่าตัวแปรทั้งสองตัวมีความสัมพันธ์กัน ในเชิงตรงกันข้ามอย่างมาก หากมีค่าใกล้ +1.0 นั้นหมายความว่า ตัวแปรทั้งสองมีความสัมพันธ์กัน โดยตรงอย่างมาก และหากมีค่าเป็น 0 นั้นหมายความว่า ตัวแปรทั้งสองตัวไม่มีความสัมพันธ์ต่อกัน รวมถึง เลือกว่าจะใช้ข้อมูลที่มีความสัมพันธ์ต่อการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวานตามทฤษฎีทางการแพทย์ เมื่อกำหนดข้อมูลผลลัพธ์ คือ eGFR2 พบว่าข้อมูลคุณลักษณะ ที่สำคัญ 3 อันดับแรกของการทำนายผลลัพธ์ ได้แก่ อัตราการกรองของไตครั้งที่ 1 ที่ค่า 0.81 อายุ คำนีวมวลกาย ไชมันชนิดดี ที่ค่า 0.15 และระยะเวลาที่ผู้ป่วยเป็นโรครมาแล้วกี่ปี ที่ค่า 0.11 ตามลำดับ

#### 4.1.3.4 ผลการหานัยสำคัญทางสถิติของข้อมูลสุขภาพ

เมื่อทำการทดสอบสมมติฐานทางสถิติของข้อมูลสุขภาพ พบว่าได้ผลค่านัยสำคัญ แสดงดังตารางที่ 4.3

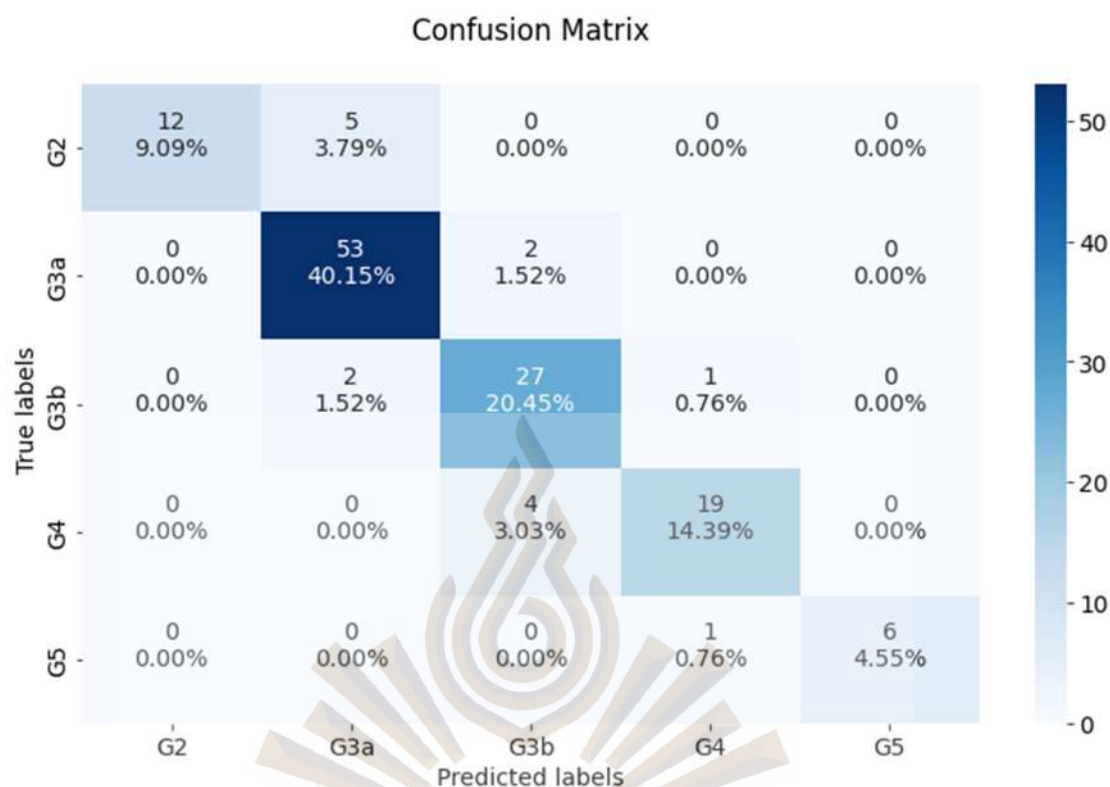
ตารางที่ 4.3 แสดงนัยสำคัญทางสถิติของข้อมูลสุขภาพ

ลำดับ	ข้อมูล	P-Value
1	eGFR1 (mi/min/1.73 m <sup>2</sup> )	0.552
2	Age (Year)	0.216
3	Duration (Year)	0.043
4	BMI_change (kg/m <sup>2</sup> )	0.017
5	SBP_change (mmHg)	0.060
6	DBP_change (mmHg)	0.113
7	FBS_change (mg/DL)	0.014
8	HbA1c_change (%)	0.207
9	CHO_change (mg/DL)	0.041
10	TG_change (mg/DL)	0.010
11	LDL_change (mg/DL)	0.028

จากตารางที่ 4.3 จากค่า P-Value มีระดับนัยสำคัญทางสถิติต่ำกว่า 0.05 หรือ 0.025 ซึ่งแสดงให้เห็นว่าผลที่ได้จากการศึกษาเพียงพอที่จะสามารถนำมาใช้ในการทดสอบระบบ นอกจากนี้แสดงถึงความสัมพันธ์ของข้อมูลว่าตัวแปรตาม (Y) มีความสัมพันธ์กับตัวแปรอิสระ (X) เมื่อกำหนดตัวแปรอิสระ และกำหนดตัวแปรตามแล้ว ทำการทดสอบค่าระดับนัยสำคัญของตัวแปรตามที่เลือกใช้ จึงใช้ข้อมูลในการเรียนรู้และทดสอบระบบนอกจากนี้มีตัวแปรอิสระที่ไม่มีความสัมพันธ์กับข้อมูลแต่ต้องนำมาใช้ในการทดสอบระบบเนื่องจากมีความสัมพันธ์กันในทางการแพทย์

4.1.3.5 ผลการวัดประสิทธิภาพแบบจำลองของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ในการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน

เมื่อทำการวิเคราะห์ข้อมูลที่น่าจะระบบ พบว่าได้ผลค่า Confusion matrix แสดงดังรูปที่ 4.16

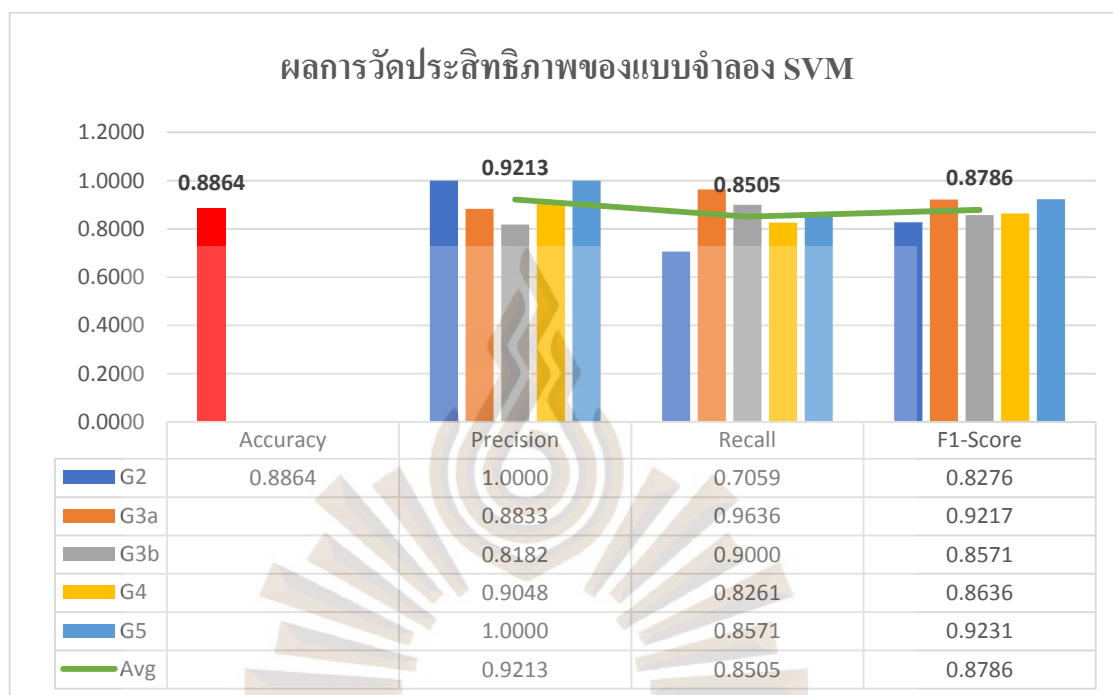


รูปที่ 4.16 แสดงค่า Confusion Matrix ของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน

จากรูปที่ 4.16 แสดงค่า Confusion Matrix เป็นตารางที่ใช้ประเมินผลลัพธ์การทำนาย เปรียบเทียบกับค่าจริง ของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน โดยแบ่ง ข้อมูลผลลัพธ์ ออกเป็น 5 กลุ่ม คือแสดงถึงผู้ป่วยเบาหวานที่เป็นโรคไตเรื้อรังในระยะ G2 มีค่าระดับอัตราการกรองของไตระหว่าง 60-89 มิลลิลิตร/นาที/1.73 ตารางเมตร ระยะ G3a มีค่าระดับอัตราการกรองของไตระหว่าง 45-59 มิลลิลิตร/นาที/1.73 ตารางเมตร ระยะ G3b มีค่าระดับอัตราการกรองของไตระหว่าง 30-44 มิลลิลิตร/นาที/1.73 ตารางเมตร ระยะ G4 มีค่าระดับอัตราการกรองของไตระหว่าง 15-29 มิลลิลิตร/นาที/1.73 ตารางเมตร ระยะ G5 มีค่าระดับอัตราการกรองของไตน้อยกว่า 15 มิลลิลิตร/นาที/1.73 ตารางเมตร จากนั้นระบบจะทำการวิเคราะห์และตรวจสอบความแม่นยำของระบบ พบว่า ระยะ G2 ทำนายได้ถูกต้อง 12 จาก 17 คน หรือเท่ากับ 9.09% ระยะ G3a ทำนายได้ถูกต้อง 53 จาก 55 คน หรือเท่ากับ 40.15% ระยะ G3b ทำนายได้ถูกต้อง 27 จาก 30 คน หรือเท่ากับ 20.45% ระยะ G4 ทำนายได้ถูกต้อง 19 จาก 23 คน หรือเท่ากับ 14.39% และระยะ G5 ทำนายได้ถูกต้อง 6 จาก 7 คน หรือเท่ากับ 4.55 %



เมื่อทำการวิเคราะห์ข้อมูลที่นำเข้าระบบพบว่าได้ผลค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก ค่าความถ่วงดุล แสดงดังรูปที่ 4.17

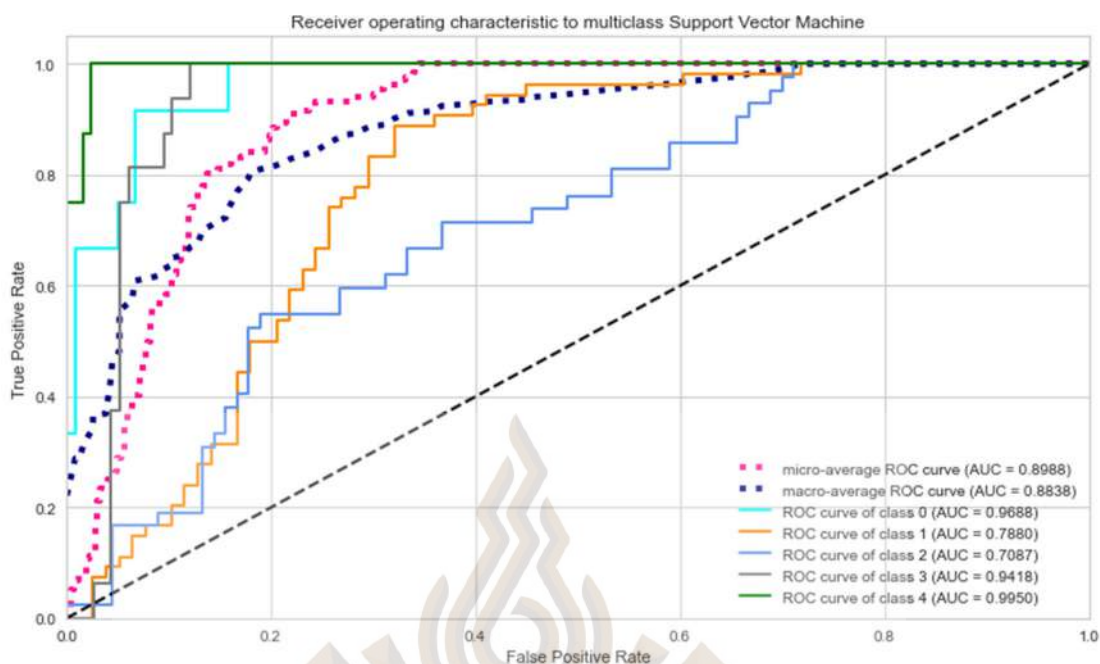


รูปที่ 4.17 กราฟแสดงผลการวัดประสิทธิภาพของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน

จากรูปที่ 4.17 แสดงผลการวัดประสิทธิภาพของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน มีค่าความถูกต้องทั้งหมด อยู่ที่ 0.8864 ค่าความแม่นยำ ระยะ G2 ให้ค่า 1.0000 ระยะ G3a ให้ค่า 0.8833 ระยะ G3b ให้ค่า 0.8182 ระยะ G4 ให้ค่า 0.9048 ระยะ G5 ให้ค่า 1.000 ค่าความระลึก ระยะ G2 ให้ค่า 0.7059 ระยะ G3a ให้ค่า 0.9636 ระยะ G3b ให้ค่า 0.9000 ระยะ G4 ให้ค่า 0.8261 ระยะ G5 ให้ค่า 0.8571 ค่าความถ่วงดุล ระยะ G2 ให้ค่า 0.8276 ระยะ G3a ให้ค่า 0.9217 ระยะ G3b ให้ค่า 0.8571 ระยะ G4 ให้ค่า 0.8636 และระยะ G5 ให้ค่า 0.9231

เมื่อเปรียบเทียบแต่ละอัลกอริทึมโดยพิจารณาจากค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ ROC Curve สามารถแสดงได้ดังรูปที่ 4.18





รูปที่ 4.18 แสดงพื้นที่ใต้กราฟ ROC Curve ของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน โดยแบ่งผลลัพธ์เป็น 6 กลุ่ม

จากรูปที่ 4.18 แสดงผลการวัดประสิทธิภาพของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ด้วยค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ โดยกราฟแสดงความสัมพันธ์ การทำนายถูกต้องไปทางแกนตั้ง (Y) และถ้าทำนายผิดไปแนวแกนนอน (X) และถ้าค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ มีค่าเข้าใกล้ 1 จะแสดงว่าข้อมูลผลลัพธ์ในกลุ่มนั้นมีประสิทธิภาพที่ดี โดย ระยะ G2 ให้ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ 0.9688 ระยะ G3a ให้ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ 0.7880 ระยะ G3b ให้ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ 0.7087 ระยะ G4 ให้ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ 0.9418 ระยะ G5 ให้ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ 0.9950

#### 4.1.3 ผลการพัฒนาเว็บแอปพลิเคชันสำหรับตรวจประเมินระยะของโรคในผู้ป่วยโรคไตเรื้อรัง โดยใช้เทคโนโลยีปัญญาประดิษฐ์

เมื่อพัฒนาเว็บแอปพลิเคชันการวิเคราะห์ระยะการเกิดโรคไตเรื้อรังและทำนายระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวานได้ดังรูปที่ 4.19



โดยใช้พื้นที่ได้กราฟ อยู่ที่ 100% สูงสุดลำดับที่ 1 เป็นค่าที่ทำให้เกิดกรณีการที่แบบจำลองมีความแม่นยำสูงมากในตอนทดลองแต่พอเอาไปใช้งานจริงกลับได้ผลลัพธ์ที่ไม่ดีของระบบได้

## 4.2 ผลการทดสอบ

ผลการทดสอบระบบวิเคราะห์และทำนายความเสี่ยงโรคไตเรื้อรังของผู้ป่วยเบาหวานแบ่งออกเป็น 3 ส่วน ดังนี้

### 4.2.1 ผลการทดสอบเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน

ผลการทดสอบเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการวิเคราะห์โอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานจะได้ค่าตัววัดประสิทธิภาพ โดยใช้พื้นที่ได้กราฟ ของระบบดังตารางที่ 4.4

ตารางที่ 4.4 แสดงผลการทดลองของโอกาสการเกิดโรคไตเรื้อรัง

เมตริกซ์ที่ใช้วัดความถูกต้องของแบบจำลอง	ผลการทดลอง
ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ได้กราฟ	97 %

จากตารางที่ 4.4 แสดงผลการทดลองของโอกาสการเกิดโรคไตเรื้อรังระบบจะแสดงผลการประมวลผลที่รวดเร็วและสามารถช่วยสนับสนุนในการตัดสินใจของแพทย์ โดยมีค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ได้กราฟ ของระบบอยู่ที่ 97 % เนื่องจากอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ซึ่งมีค่าสูงสุดลำดับที่ 2 จากการเปรียบเทียบประสิทธิภาพของอัลกอริทึมทั้งหมด 4 อัลกอริทึม เนื่องจากอัลกอริทึมที่ให้ค่าตัววัดประสิทธิภาพโดยใช้พื้นที่ได้กราฟ อยู่ที่ 100% สูงสุดลำดับที่ 1 เป็นค่าที่ทำให้เกิดกรณีการที่แบบจำลองมีความแม่นยำสูงมากในตอนทดลองแต่พอเอาไปใช้งานจริงกลับได้ผลลัพธ์ที่ไม่ดีของระบบได้ โดยระบบสามารถแสดงผลโอกาสการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานว่ามีความเสี่ยงเป็นโรคไตเรื้อรังหรือไม่

#### 4.2.2 ผลการทดสอบการวิเคราะห์ระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน



รูปที่ 4.20 แสดงผลการวิเคราะห์โรคไตเรื้อรังสำหรับผู้ป่วยเบาหวาน

จากรูปที่ 4.20 ผลการวิเคราะห์สามารถบอกได้ว่าผู้ป่วยเบาหวานมีโอกาสการเกิดโรคไตเรื้อรังหรือไม่และผู้ป่วยมีโอกาสเกิดโรคไตเรื้อรังได้ในระยะใด โดยทดสอบจากการใส่ค่าดังต่อไปนี้ได้แก่

1) อัตราการกรองของไต	เท่ากับ	44
2) อายุ (ปี)	เท่ากับ	79
3) ระยะเวลาที่ผู้ป่วยเป็นโรคมานี้กี่ปี (ปี)	เท่ากับ	18
4) ดัชนีมวลกาย	เท่ากับ	34.6
5) ความดันสูงสุดของหัวใจห้องล่างบีบตัว	เท่ากับ	205
6) ไ้ไขมันชนิดดี	เท่ากับ	36
7) ความดันเลือดที่ต่ำที่สุดขนาดหัวใจห้องล่างคลายตัว	เท่ากับ	95
8) การตรวจระดับน้ำตาลในเลือดหลังจากอดอาหาร 8 ชั่วโมง	เท่ากับ	88
9) ฮีโมโกลบิน เอ วัน ซี	เท่ากับ	6
10) คอเลสเตอรอล	เท่ากับ	109
11) ไ้ไตรกลีเซอไรด์	เท่ากับ	99
12) ไ้ไขมันชนิดความหนาแน่นต่ำ	เท่ากับ	53

ซึ่งผลลัพธ์ของการวิเคราะห์ระยะการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน เท่ากับ G3b Moderately to Severely Decreased ซึ่งหมายถึงเป็นระยะที่ไตทำงานผิดปกติปานกลาง ระดับที่ 2 ถูกต้องตั้งมาตรฐานของระยะการเกิดโรคไตเรื้อรัง

#### 4.2.3 ผลทดสอบการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน





5) ความดันสูงสุดของหัวใจห้องล่างบีบตัว	เท่ากับ	105
6) ไชมันชนิดดี	เท่ากับ	28
7) ความดันเลือดที่ต่ำที่สุดขนาดหัวใจห้องล่างคลายตัว	เท่ากับ	81
8) การตรวจระดับน้ำตาลในเลือดหลังจากอดอาหาร 8 ชั่วโมง	เท่ากับ	138
9) ฮีโมโกลบิน เอ วัน ซี	เท่ากับ	7
10) คอเลสเตอรอล	เท่ากับ	121
11) ไตรกลีเซอไรด์	เท่ากับ	302
12) ไชมันชนิดความหนาแน่นต่ำ	เท่ากับ	111

ซึ่งผลลัพธ์ของการทำนายการเกิดโรคไตเรื้อรังและระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวาน เท่ากับ G4 Severely Decreased ซึ่งหมายถึงเป็นระยะที่ไตทำงานผิดปกติอย่างมาก ถูกต้องตั้งมาตรฐานของระยะการเกิดโรคไตเรื้อรัง

#### 4.2.4 ผลการทดสอบสมรรถนะเว็บแอปพลิเคชัน

ตารางที่ 4.5 ผลการทดสอบสมรรถนะของระบบ

หัวข้อการทดสอบ	ผลการทดสอบ
1) ความถูกต้องของระบบ	ผ่าน
2) ความน่าเชื่อถือของระบบ	ผ่าน
3) ความแม่นยำของระบบ	ผ่าน

#### 4.2.5 ผลการทดสอบความปลอดภัยของการวิจัย

ระบบมีความปลอดภัยและข้อมูลของผู้ป่วยถูกจัดเก็บไว้ในเซิร์ฟเวอร์ของวิทยาลัยวิศวกรรมชีวการแพทย์ มหาวิทยาลัยรังสิต โดยผู้ที่สามารถเข้าถึงข้อมูลได้นั้นจะต้องเป็นผู้วิจัยเท่านั้นโดยมีการตรวจสอบความปลอดภัยก่อนเข้าห้องทดสอบ และจำกัดสิทธิการเข้าถึงข้อมูลด้วยการกรอกรหัสผ่าน ดังตารางที่ 4.6

ตารางที่ 4.6 ผลการทดสอบความปลอดภัยของการวิจัย

หัวข้อการทดสอบ	ผลการทดสอบ
1) การจำกัดสิทธิการเข้าถึงข้อมูล ของผู้ใช้งานในแต่ละระดับ	ผ่าน
2) การตรวจสอบความปลอดภัยก่อนเข้าห้องทดสอบ	ผ่าน

จากตารางที่ 4.6 ผลการทดสอบความปลอดภัยของการวิจัยทางทีมผู้วิจัยได้มีการจำกัดสิทธิการเข้าถึงข้อมูล โดยการกรอกรหัสก่อนเข้าใช้งานคอมพิวเตอร์ในส่วนของจัดการข้อมูล และทำการทดสอบระบบ รวมไปถึงมีการตรวจสอบความปลอดภัยก่อนเข้าห้องทดสอบด้วยการสแกนลายนิ้วมือ

#### 4.2.6 ผลการประเมินเว็บแอปพลิเคชัน

ผลการประเมินเว็บแอปพลิเคชัน โดยผู้ใช้งาน คือ ผู้เชี่ยวชาญทางการแพทย์และผู้เกี่ยวข้อง จำนวน 6 คน ตามหัวข้อในการประเมิน 6 ด้านแสดงดังตารางที่ 4.7

ตารางที่ 4.7 แสดงผลการประเมินเฉลี่ยหลังการทดสอบการใช้งานเว็บแอปพลิเคชัน

รายการประเมิน	ค่าเฉลี่ย	แปลผล
1) ด้านความสวยงามของเว็บแอปพลิเคชัน	4.43	มาก
2) ด้านการอำนวยความสะดวกภายในเว็บแอปพลิเคชัน	4.40	มาก
3) ด้านประโยชน์จากการทำนายโรคของเว็บแอปพลิเคชัน	4.50	มาก
4) ด้านความรวดเร็วในการตอบสนองการใช้งาน	4.53	มากที่สุด
ความชัดเจนของข้อมูล	4.43	มาก
6) ด้านความพึงพอใจในการใช้งาน	4.53	มากที่สุด
คะแนนเฉลี่ยรวม	4.47	มาก

## บทที่ 5

### สรุปผลและข้อเสนอแนะ

จากผลการศึกษาเรื่องการสร้างระบบวิเคราะห์และทำนายความเสี่ยงการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน สามารถสรุปผลโดยแบ่งออกเป็น 3 ส่วน ดังนี้

- 1) สรุปผลการจัดทำวิจัย
- 2) การอภิปรายผลการจัดทำวิจัย
- 3) ข้อเสนอแนะ

#### 5.1 สรุปผลการจัดทำวิจัย

การสร้างระบบวิเคราะห์และทำนายความเสี่ยงการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานมีวัตถุประสงค์เพื่อวิเคราะห์และทำนายความเสี่ยงผู้ป่วยเบาหวานเป็นโรคไตเรื้อรังหรือไม่ และทำนายโอกาสเกิดโรคไตเรื้อรังในระยะต่อไป โดยระบบจะสามารถนำมาประยุกต์ใช้การกับโรคต่าง ๆ ได้ และผลลัพธ์ที่ได้ออกมาจะช่วยสนับสนุนการตัดสินใจของแพทย์ให้มีความแม่นยำมากขึ้น นอกจากนี้แพทย์ยังสามารถตรวจสอบพฤติกรรมของผู้ป่วยได้จากผลการวิเคราะห์ข้อมูลของระบบ รวมไปถึงระบบสามารถเพิ่มข้อมูลของผู้ป่วยรายใหม่ได้ โดยหลังจากที่เพิ่มข้อมูลใหม่แล้วระบบจะทำการวิเคราะห์และทำนายผลให้กับผู้ป่วยรายนี้ทำให้แพทย์สามารถเตรียมการวินิจฉัยและการรักษาได้ล่วงหน้าอย่างทันท่วงที

ในการศึกษาการออกแบบและการสร้างระบบวิเคราะห์และทำนายข้อมูล ระบบที่สร้างขึ้นจะต้องมีพื้นที่เพียงพอต่อการรองรับข้อมูลจำนวนมาก ระบบจะต้องมีความถูกต้องและความแม่นยำในการวิเคราะห์ข้อมูลเป็นการศึกษาเพื่อให้เห็นภาพรวมทั้งหมด ภายใต้การทำงานของระบบ ออกเป็น 4 ส่วน คือ ส่วนที่ 1 การนำข้อมูลของผู้ป่วยโรคไตเรื้อรังและผู้ป่วยโรคเบาหวาน ซึ่งจะประกอบด้วย ผู้ป่วยเบาหวานที่ไม่เป็นโรคไตเรื้อรังและผู้ป่วยเบาหวานที่เป็นโรคไตเรื้อรัง เพื่อวิเคราะห์หาอัลกอริทึมที่เหมาะสม ส่วนที่ 2 พัฒนาระบบด้วยกระบวนการออกแบบระบบพื้นฐานโดยภาษาไพธอน ส่วนที่ 3 การทดสอบเปรียบเทียบประสิทธิภาพของแบบจำลอง การประเมิน

ความสามารถของแบบจำลอง ส่วนที่ 4 เป็นการแสดงผลการทดสอบของระบบ ได้แก่ ผู้ป่วยเบาหวานจะเป็นโรคไตเรื้อรังหรือไม่ และผู้ป่วยเบาหวานที่เป็นโรคไตเรื้อรังสามารถมีโอกาสที่จะเกิดในระยะต่อไป โดยค่าที่จะแสดงผลจะเป็นเปอร์เซ็นต์ความเสี่ยงที่ได้จากการวิเคราะห์ข้อมูล

จากการทดสอบการการสร้างระบบวิเคราะห์และทำนายความเสี่ยงการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน โดยใช้อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน พบว่าผลการทำงานของระบบเป็นไปตามวัตถุประสงค์ที่ออกแบบไว้ข้างต้น รวมถึงระบบมีความเป็นอัจฉริยะมากขึ้น นอกจากนี้ระบบสามารถช่วยวิเคราะห์และทำนายผลของข้อมูล โดยระบบจะแสดงผลการประมวลผลที่รวดเร็ว และสามารถช่วยสนับสนุนในการตัดสินใจของแพทย์ โดยมีค่าความแม่นยำ ของระบบอยู่ที่ 97 % ในการวิเคราะห์ว่าผู้ป่วยเบาหวานที่มีความเสี่ยงเป็นโรคไตเรื้อรังหรือไม่ โดยระบบวิเคราะห์และทำนายความเสี่ยงการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานได้รับการประเมินจากแพทย์ผู้ใช้งานว่าระบบที่จัดทำนั้นมีประสิทธิภาพและสามารถนำไปประยุกต์ใช้งานได้จริง

ส่วนการพัฒนาเว็บแอปพลิเคชันสำหรับการวิเคราะห์ระยะการเกิดโรคไตเรื้อรังและทำนายระยะของโรคไตเรื้อรังในอนาคตของผู้ป่วยเบาหวานผ่านเครื่องคอมพิวเตอร์หรือสมาร์ตโฟน โดยนำแบบจำลองจากอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ซึ่ง แบบจำลองที่ได้นั้นเหมาะสมที่สุดจากการทดสอบประสิทธิภาพของทั้ง 4 อัลกอริทึม เข้ามาช่วยในการวิเคราะห์ข้อมูลและประมวลผลพร้อมแสดงผลลัพธ์ระยะของโรคไตเรื้อรังในผู้ป่วยเบาหวาน พัฒนาระบบด้วยโปรแกรม Project jupyter และภาษาไพธอน ในการทดสอบระบบให้ผู้ใช้ งาน คือ ผู้เชี่ยวชาญทางการแพทย์และผู้เกี่ยวข้อง จำนวน 6 คน มีหัวข้อในการประเมิน 6 ด้าน ได้แก่ 1) ด้านความสวยงามของเว็บแอปพลิเคชัน ค่าเฉลี่ยเท่ากับ 4.33 อยู่ในระดับมาก 2) ด้านการอำนวยความสะดวกภายในเว็บแอปพลิเคชัน ค่าเฉลี่ยเท่ากับ 4.00 อยู่ในระดับมาก 3) ด้านประโยชน์จากการทำนายโรคของเว็บแอปพลิเคชัน ค่าเฉลี่ยเท่ากับ 4.83 อยู่ในระดับมากที่สุด 4) ด้านความรวดเร็วในการตอบสนองการใช้งาน ค่าเฉลี่ยเท่ากับ 4.33 อยู่ในระดับมาก 5) ด้านความชัดเจนของข้อมูล ค่าเฉลี่ยเท่ากับ 4.83 อยู่ในระดับมากที่สุด 6) ด้านความพึงพอใจในการใช้งาน ค่าเฉลี่ยเท่ากับ 4.50 อยู่ในระดับมากที่สุด มีค่าเฉลี่ยในภาพรวมเท่ากับ 4.47 อยู่ในระดับมาก



## 5.2 อภิปรายผลการจัดทำวิจัย

การจัดทำวิจัยระบบวิเคราะห์และทำนายความเสี่ยงการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานโดยใช้หน่วยประมวลผลของคอมพิวเตอร์GPU ซึ่งเป็นสิ่งจำเป็นต่อการเรนเดอร์ข้อมูลจำนวนมาก และการใช้ภาษาไพธอน ในการพัฒนาระบบ รวมไปถึงการศึกษาแบบจำลอง ที่ใช้ในการหาความสัมพันธ์ของข้อมูล โดยเลือกอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน เพื่อใช้ในการสร้างระบบวิเคราะห์และทำนายความเสี่ยงการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวานเพื่อนำไปใช้ร่วมกับคณะแพทยศาสตร์วชิรพยาบาล มหาวิทยาลัยนวมินทราธิราช ระบบที่สร้างขึ้นสามารถเพิ่มข้อมูลของผู้ป่วยรายใหม่ได้ อีกทั้งระบบยังสามารถนำไปพัฒนาให้ใช้ร่วมกับโรคอื่น ๆ ได้ด้วย

## 5.3 ข้อเสนอแนะ

งานวิจัยนี้เป็นการนำเสนอผลการประยุกต์ใช้ปัญญาประดิษฐ์ทำนายความเสี่ยงการเกิดโรคไตเรื้อรังของผู้ป่วยเบาหวาน โดยใช้ข้อมูลทางคลินิกของผู้ป่วยซึ่งข้อมูลดังกล่าวยังสามารถใช้ทำนายโรคอื่น ๆ ได้อีกหากจัดระบบข้อมูลให้ดีจะเพิ่มความสามารถในการทำนายโรคของผู้ป่วยได้เพิ่มขึ้น หรืออาจทำนายโรคอื่น ๆ ด้วยในคราวเดียวกัน

## บรรณานุกรม

- โรงพยาบาลบางปะกอก 9 อินเตอร์เนชั่นแนล. (2560). *วิธีคำนวณดัชนีมวลกายBMI*. สืบค้นจาก [https://www.bangpakokhospital.com/care\\_blog/content/วิธีคำนวณดัชนีมวลกายBMI](https://www.bangpakokhospital.com/care_blog/content/วิธีคำนวณดัชนีมวลกายBMI).
- โรงพยาบาลสถาบันโรคไตภูมิราชนครินทร์. (ม.ป.ป.). [http://www.rajavithi.go.th/rj/wp-content/uploads/2017/05/2.BKI\\_.pdf](http://www.rajavithi.go.th/rj/wp-content/uploads/2017/05/2.BKI_.pdf).
- กษัตริย์ วิชาสวรัช, กมล โขมิตรังสิกุลม, กฤษณพงศ์ มโนธรรม, การยोजना ตั้งนรราชชกิจ, กำร สีสามะลิ, กิตติ์วิ กฤษณ์เมธาภาคย์ม และอังคณีย์ ชนะกุล. (2562). *ตำราโรคไตเรื้อรัง (พิมพ์ครั้งที่ 2)*. กรุงเทพฯ: เท็กซ์ แอนด์ เจอร์นัลส์ พับลิเคชัน จำกัด
- ฉลอง สีแก้วสีว.(2012). *ความรู้เบื้องต้นเกี่ยวกับสถิติ (Basic Statistics)*. สืบค้นจาก <https://sites.google.com/site/mystatistics01/chapter1/basic-theory>
- ฉลอง สีแก้วสีว.(2012). *สมการการถดถอย Analysis คืออะไร*. สืบค้นจาก <https://sites.google.com/site/mystatistics01/สมการการถดถอย-correlation-analysis/สมการการถดถอย-Analysis>
- ปิยะธิดา จึงสมาน. (ม.ป.ป.). *โรคไตเรื้อรัง(Chronic Kidney disease) โจทย์ใหม่ท้าทายคนทำงาน ควบคุมโรค*. สืบค้นจาก <http://irem.ddc.moph.go.th/uploads/tiny/conference/2559/docs/after/ppt/4-CKD.pdf>
- ประพนธ์ ตั้งศรีเกียรติกุล. (2020). *รวมพลัง “คัดกรอง ป้องกัน รู้ทัน โรคไต”*. สืบค้นจาก <https://www.ryt9.com/s/prg/3096140>
- Achakulvisut, T. (2018). *[ML] Bagging หรือ Boosting คืออะไร ทำงานอย่างไร?*. สืบค้นจาก <https://tupleblog.github.io/bagging-boosting/>
- Arnondora. (2019). *Feature Selection Machine Learning*. Retrieved from <https://arnondora.in.th/feature-selection-machine-learning>
- Chalermkiatsakul, P. (2018). *Supervised Learning คืออะไร? ทำงานยังไง?*. Retrieved from <https://phuri.medium.com/supervised-learning>
- Chuengsaman, P. (n.d.). *โรคไตเรื้อรัง(Chronic Kidney disease) โจทย์ใหม่ท้าทายคนทำงานควบคุมโรค*. Retrieved from <http://irem.ddc.moph.go.th/uploads/tiny/conference/2559/docs/after/ppt/4-CKD.pdf>
- Echouffo-Tcheugui, J. B., & Kengne, A. P. (2012). Risk models to predict chronic kidney disease and its progression: a systematic review. *PLoS medicine*, 9(11), e1001344. Retrieved from <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001344>.

## บรรณานุกรม (ต่อ)

- Kumar, A. (2019). *Difference between Machine Learning & Traditional Software*. Retrieved from <https://vitalflux.com/difference-between-machine-learning-traditional-software/>
- Latysheva, N. (2016). *Implementing Your Own k-Nearest Neighbor Algorithm Using Python*. Retrieved from <https://www.kdnuggets.com/2016/01/implementing-your-own-knn-using-python.html>
- Mohajon, J. (2020). *Confusion Matrix for Your Multi-Class Machine Learning Model*. Retrieved from <https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826>
- Navlani, A. (2018). *KNN Classification using Scikit-learn*. Retrieved from <https://www.datacamp.com/tutorial/k-nearest-neighbor-classification-scikit-learn>
- Navlani, A. (2018). *Support Vector Machines with Scikit-learn*. Retrieved from <https://www.datacamp.com/tutorial/svm-classification-scikit-learn-python#svm>
- Pathak, M. (2019). *Using XGBoost in Python*. Retrieved from <https://www.datacamp.com/community/tutorials/xgboost-in-python>
- Samad, A. (2019). *What is Machine Learning*. Retrieved from <https://becominghuman.ai/what-is-machine-learning-d292114cc6ce>
- Singh, A. (2018). *A Comprehensive Guide to Ensemble Learning (with Python codes)*. Retrieved from <https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/>
- Tangri, N., Stevens, L. A., Griffith, J., Tighiouart, H., Djurdjev, O., Naimark, D., ... & Levey, A. S. (2011). A predictive model for progression of chronic kidney disease to kidney failure. *Jama*, 305(15), 1553-1559. Retrieved from <https://towardsdatascience.com/การถดถอย-or-classification-linear-or-logistic-f093e8757b9c>
- Vijite, P. (2018). *ประเภทของ Machine Learning*. Retrieved from <https://medium.com/coeffest>
- Xiao, J., Ding, R., Xu, X., Guan, H., Feng, X., Sun, T., ... & Ye, Z. (2019). Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. *Journal of translational medicine*, 17(1), 1-13. Retrieved from <https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-019-1860-0>





ภาคผนวก ก

ตัวอย่างแบบประเมินความพึงพอใจ

มหาวิทยาลัยรังสิต Rangsit University



## แบบประเมินงานวิจัย

ชื่อเรื่อง เว็บแอปพลิเคชันสำหรับตรวจประเมินระยะของโรคในผู้ป่วยโรคไตเรื้อรัง โดยใช้เทคโนโลยีปัญญาประดิษฐ์  
ชื่อผู้วิจัย นางสาวสุกานา ช่างบุญ

ลิงค์เว็บแอปพลิเคชัน <https://share.streamlit.io/walaiwat/ckd-xgboost/main/app.py>

ชื่อ-นามสกุล ผู้ประเมิน \*

ข้อความคำตอบสั้นๆ

ตำแหน่ง/วิชาชีพ

ข้อความคำตอบสั้นๆ

1) ด้านความสวยงามของเว็บแอปพลิเคชัน

- มากที่สุด
- มาก
- ปานกลาง
- น้อย
- น้อยที่สุด

มหาวิทยาลัยรังสิต Rangsit University

...

2) ด้านการอ่านความสะดวกภายในเว็บแอปพลิเคชัน

- มากที่สุด
- มาก
- ปานกลาง
- น้อย
- น้อยที่สุด

3) ด้านประโยชน์จากการทำนายโรคของเว็บแอปพลิเคชัน

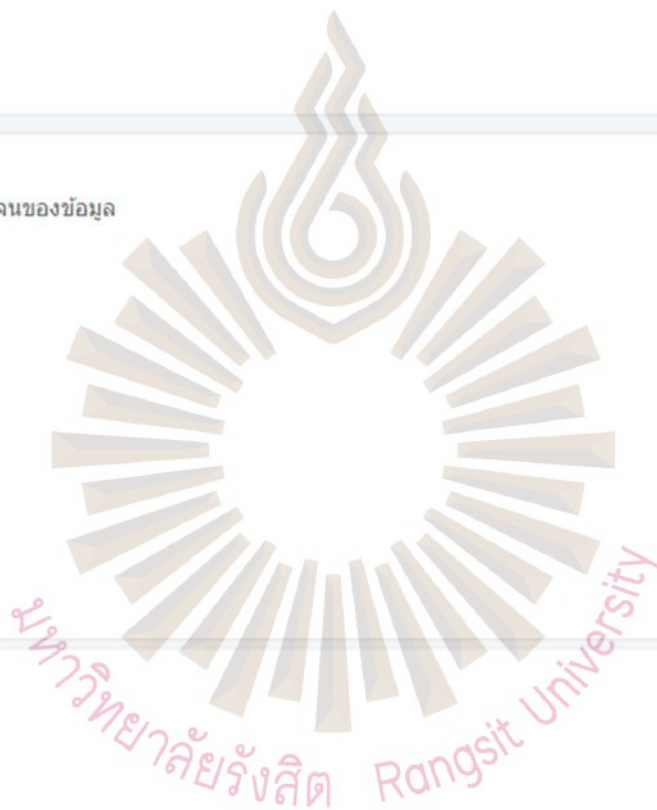
- มากที่สุด
- มาก
- ปานกลาง
- น้อย
- น้อยที่สุด

4) ด้านความรวดเร็วในการตอบสนองการใช้งาน

- มากที่สุด
- มาก
- ปานกลาง
- น้อย
- น้อยที่สุด

5) ด้านความชัดเจนของข้อมูล

- มากที่สุด
- มาก
- ปานกลาง
- น้อย
- น้อยที่สุด





ภาคผนวก ข

การนำเสนอผลงานในงานประชุมวิชาการ

**The 17 UTCC National Graduate Research Conference**

มหาวิทยาลัยรังสิต Rangsit University



UTCC



การประชุมนำเสนอผลงานวิจัยบัณฑิตศึกษาระดับชาติ ครั้งที่ 17 ปีการศึกษา 2565  
*The 17<sup>th</sup> UTCC National Graduate Research Conference 2022*

ขออบประกาศเกียรติคุณฉบับนี้ เพื่อแสดงว่า

สฤณา ช่างมู๋

ได้นำเสนอผลงานในหัวข้อ

เว็บแอปพลิเคชันสำหรับตรวจสอบประเมินระยะของโรคในผู้ป่วยโรคไตเรื้อรัง  
โดยใช้เทคโนโลยีปัญญาประดิษฐ์

ณ บัณฑิตวิทยาลัย มหาวิทยาลัยหอการค้าไทย

ให้ไว้ ณ วันที่ 10 สิงหาคม พ.ศ. 2565

(ผู้ช่วยศาสตราจารย์ ดร. มานา บัจฉิมนันท์)

คณบดีบัณฑิตวิทยาลัย มหาวิทยาลัยหอการค้าไทย





Cisco Webex Meetings Meeting Info Hide Menu Bar ^

45:49

File Edit Share View Audio & Video Participant Meeting Help

Viewing sakuna Changboo's application... 53% +

Speaking: sakuna Changboo

### WEB APPLICATION โดยใช้ AI

**Prediction for CKD stage**

How old are you :	81
How many year have you had the disease ?	28
What is your BMI :	27.08
What is your SBP :	180
What is your DBP :	95
What is your FBS :	122
What is your A1c :	7.7
What is your CHO :	181
What is your TG :	176
What is your LDL :	79
GFR Stage :	G4 Severely decreased

ลด Human error

แม่นยำ รวดเร็ว

ลดอัตราการวินิจฉัยโรคผิดพลาด

AI

Participants Chat

Type here to search 29°C 10:22 AM 8/10/2022



## ประวัติผู้วิจัย

ชื่อ	สกุณา ช่างบุญ
วัน เดือน ปีเกิด	16 เมษายน 2533
สถานที่เกิด	จังหวัดอ่างทอง ประเทศไทย
ประวัติการศึกษา	มหาวิทยาลัยกรุงเทพ ปริญญาบริหารธุรกิจบัณฑิต สาขาวิชาคอมพิวเตอร์ธุรกิจ, 2554 มหาวิทยาลัยรังสิต ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมชีวการแพทย์, 2565
ที่อยู่ปัจจุบัน	117 โรงแรมธาราเพลส ถนนสามเสน แขวงวัดสามพระยา เขตพระนคร กรุงเทพมหานคร 10200
สถานที่ทำงาน	สหกรณ์ออมทรัพย์พนักงานธนาคารแห่งประเทศไทย จำกัด
ตำแหน่งปัจจุบัน	เจ้าหน้าที่วิเคราะห์และจัดการข้อมูล