

การพัฒนาระบบการคัดกรองกลุ่มเสี่ยงโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูล

พวงทอง วงอนันต์¹ และ อีรพงษ์ สังข์ศรี²

^{1,2}คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครราชสีมา นครราชสีมา

Emails: mabmanza2@windowslive.com, evan.s.nrru@hotmail.com

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบการคัดกรองกลุ่มเสี่ยงโรคเบาหวานในรูปแบบของเว็บแอปพลิเคชัน (Web-based Application) โดยการประยุกต์ใช้เทคนิคทางด้านเหมืองข้อมูลเข้ามา การดำเนินงานจะเป็นการศึกษาและวิเคราะห์ปัจจัยที่มีผลต่อการเกิดโรคเบาหวานเพื่อนำมาสร้างเป็นตัวแบบการพยากรณ์การเกิดโรคเบาหวานด้วยเทคนิคต้นไม้ตัดสินใจ (Decision Tree Technique) และการพัฒนาระบบการคัดกรองกลุ่มเสี่ยงโรคเบาหวานด้วยตัวแบบการพยากรณ์การเกิดโรคเบาหวานด้วยภาษา PHP ข้อมูลที่ใช้เป็นข้อมูลการคัดกรองโรคเบาหวานของโรงพยาบาลส่งเสริมสุขภาพตำบลละหานทราย 1,000 ชุดข้อมูล ผลจากการทดลองพบว่าประสิทธิภาพด้านความถูกต้องของตัวแบบพยากรณ์การเกิดโรคเบาหวานที่ระบบได้สร้างขึ้นมาด้วยการแบ่งแยกร้อยละของชุดข้อมูลแบบ 80:20 มีค่าเท่ากับร้อยละ 74.5 และผลประเมินความพึงพอใจในการใช้งานระบบการคัดกรองโรคเบาหวานแบบออนไลน์เฉลี่ยเท่ากับ 3.09 (ระดับพอใช้) ซึ่งด้านการช่วยเพิ่มความรวดเร็วในการทำงานมีค่าคะแนนความพึงพอใจสูงสุดเท่ากับ 3.66

ABSTRACT

This research aims to develop a diabetes screening system inform a web application using a data mining technique. This research studied and analyzed the risk factors for diabetes mellitus diagnosis. The developed system is web-based application using HTML and PHP. The Investigator is assigned to the ratio between the data for Training Data and Testing Data is Percent of 80 and 20 from the entire data set the 1,000. The value of Correctly Classified Instances with a rate of 74.5%. The overall result of the satisfaction rating of this system is

at a fair level (3.09). The results of the satisfaction rating in a working performance is at a good level (3.66).

คำสำคัญ— โรคเบาหวาน; เหมืองข้อมูล; ต้นไม้ตัดสินใจ

1. บทนำ

แบบแผนการดำเนินชีวิตที่เปลี่ยนแปลงไปนำไปสู่ปัญหาน้ำหนักเกินและโรคอ้วน ซึ่งปัจจุบันถือว่าเป็นปัจจัยอันดับต้นๆของโรคไม่ติดต่อเรื้อรัง ซึ่งโรคเหล่านี้กำลังระบาดไปทั่วโลกและเพิ่มความเสี่ยงมากขึ้นเรื่อยๆ ถ้าโรคเหล่านี้ไม่ได้รับการจัดการอย่างเพียงพอจะทำให้อัตราการเสียชีวิตและคุณภาพของประชากรเพิ่มขึ้น ตลอดจนเพิ่มภาระด้านการเงิน และการสูญเสียทางเศรษฐกิจของประเทศ ประเทศไทยได้ประสบกับปัญหาเหล่านี้และกำลังพยายามที่จะจัดการกับการคุกคามของโรคไม่ติดต่อเรื้อรังโดยเฉพาะอย่างยิ่งโรคเบาหวานหรือเรียกย่อว่า โรคดีเอ็ม (DM) ซึ่งเป็นโรคเรื้อรังที่รักษาไม่หาย เป็นโรคเรื้อรังที่จัดว่ารุนแรงซึ่งความเสี่ยงของโรคขึ้นกับผลของการควบคุมโรคได้คือการควบคุมน้ำตาลในเลือดให้อยู่ในเกณฑ์ปกติ หรือใกล้เคียงเกณฑ์ปกติที่สุด ปัจจุบันเนื่องจากจำนวนผู้สูงอายุมีมากขึ้น ทั่วโลกจึงพบเบาหวานสูงขึ้นเรื่อยๆ คาดว่าจำนวนผู้ป่วยเบาหวานทั่วโลกจะเพิ่มขึ้นจาก (2554 .ศ.พ) 2011 .ศ.ล้านคนในปี ค 366 เป็น552 ล้านคนในปีค.ศ. นอกจากนี้จาก (2573 .ศ.พ) 2030 ข้อมูลของสมาพันธ์เบาหวานโลก พบว่าโรคเบาหวานยังเป็นสาเหตุสำคัญอันดับ3 ของการเสียชีวิตในผู้หญิงทุกอายุและเป็นสาเหตุสำคัญอันดับ ปี การเสียชีวิต 50 ในผู้หญิงอายุมากกว่า 1 จากโรคเบาหวานในผู้ชายมีความสำคัญรองจากอุบัติเหตุ การจราจรและโรคเอดส์ ในแต่ละวันจะมีประชากรไทยจำนวน 180 รายเสียชีวิตจากโรคเบาหวาน หรือเกือบ 8 รายต่อชั่วโมง การแก้ปัญหาเบื้องต้นด้วยการวิเคราะห์ปัจจัยเสี่ยงในผู้ป่วยโรคเบาหวานได้จากการนำข้อมูลพื้นฐานรายบุคคลมาเป็นเครื่องมือสร้างตัวแบบประเมินความเสี่ยง ซึ่งเป็นเครื่องมือที่ใช้ได้ง่ายในระดับปฐมภูมิและในประชากรทั่วไปในการประเมินตนเอง เมื่อทราบปัจจัยที่ทำให้เกิดความเสี่ยงจะช่วยกระตุ้นให้ผู้ใช้

ที่มีความเสี่ยงสูงเปลี่ยนแปลงพฤติกรรมได้อย่างตรงเป้าหมาย และส่งผลต่อการลดโอกาสเสี่ยงต่อการเกิดโรคไม่ติดต่อเรื้อรังที่อันตรายอื่นๆ แต่ในปัจจุบันประชาชนไทย อายุมากกว่า 35 ปี ร้อยละ 50 ที่เสี่ยงต่อการเป็นโรคเบาหวานเกิดจากไม่ได้รับการวินิจฉัย และไม่ทราบว่าตัวเองเป็นกลุ่มเสี่ยง จึงสูญเสียโอกาสในการดำเนินการป้องกันและรักษา รวมทั้งระบบคัดกรองและดูแลกลุ่มเสี่ยงของกระทรวงสาธารณสุขที่ยังไม่สามารถทำงานได้อย่างมีประสิทธิภาพ เนื่องจากข้อจำกัดทั้งเรื่องกำลังคน และงบประมาณ จึงทำให้ผู้ป่วยเพิ่มมากขึ้นส่งผลให้ค่าใช้จ่ายด้านสุขภาพของทั้งระดับบุคคล ครอบครัว สังคม และระดับประเทศเพิ่มขึ้น

ดังนั้นจากปัญหาดังกล่าว งานวิจัยชิ้นนี้จึงมีวัตถุประสงค์เพื่อพัฒนาระบบสารสนเทศสำหรับการคัดกรองกลุ่มเสี่ยงโรคเบาหวานโดยทำการศึกษาและวิเคราะห์ข้อมูลปัจจัยเสี่ยงในการเกิดโรคเบาหวานเพื่อสร้างเป็นตัวแบบการพยากรณ์การเกิดโรคเบาหวานด้วยเทคนิคต้นไม้ตัดสินใจซึ่งจะสามารถนำระบบดังกล่าวไปคัดกรองกลุ่มเสี่ยงการโรคเบาหวานโดยไม่ต้องอาศัยการตรวจเลือดได้ รวมทั้งเป็นเครื่องมือในการประเมินตนเอง และสามารถพัฒนาเป็นระบบผู้เชี่ยวชาญสำหรับการวินิจฉัยโรคเบาหวานต่อไปได้

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยชิ้นนี้ได้มีการศึกษาในส่วนของทฤษฎีและงานวิจัยที่เกี่ยวข้อง ดังนี้

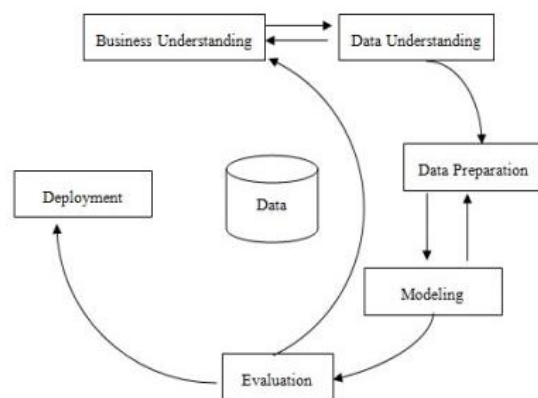
2.1. โรคไม่ติดต่อเรื้อรัง

กลุ่มโรค NCDs (Non-Communicable diseases) หรือ ชื่อภาษาไทยเรียกว่า กลุ่มโรคไม่ติดต่อเรื้อรัง นั้นเป็นชื่อเรียก กลุ่มโรคที่ไม่ได้มีสาเหตุมาจากการติดเชื้อ ไม่ได้เกิดจากเชื้อโรค ไม่สามารถติดต่อได้ผ่านการสัมผัส คลุกคลี หรือ ติดต่อกันผ่านตัวนำโรค (พาหะ) หรือสารคัดหลั่งต่างๆ หากแต่เกิดจากปัจจัยต่างๆ ภายในร่างกาย ซึ่งส่วนใหญ่เป็นผลจากไลฟ์สไตล์การใช้ชีวิต ที่มีพฤติกรรมเสี่ยงอย่าง เหล้า บุหรี่ ขาดการออกกำลังกาย อาหารหวานมันเค็มจัด และมีความเครียด ซึ่งโรคในกลุ่ม โรค NCDs ที่มีอัตราผู้ป่วยและผู้เสียชีวิตสูงสุด 6 โรค ได้แก่ โรคเบาหวาน (Diabetes Mellitus) โรคหลอดเลือดสมองและหัวใจ (Cardiovascular & Cerebrovascular Diseases) โรคถุงลมโป่งพอง (Emphysema) โรคมะเร็ง (Cancer) โรคความดันโลหิตสูง (Hypertension) โรคอ้วนลงพุง (Obesity)

2.2. เหมืองข้อมูล (Data Mining) [1]

เหมืองข้อมูล (Data Mining) เป็นการวิเคราะห์ข้อมูล และค้นหารูปแบบความสัมพันธ์ของข้อมูลที่ซ่อนเร้นอยู่โดย

วิธีการอัตโนมัติหรือกึ่งอัตโนมัติ จุดประสงค์หลักของการทำเหมืองข้อมูลคือ การดึงเอารูปแบบความสัมพันธ์ของข้อมูลออกมาเพื่อเพิ่มคุณค่าภายในของตัวข้อมูลและสร้างเป็นองค์ความรู้ใหม่ขึ้นมา การวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูลมีกระบวนการมาตรฐานที่เรียกว่า “CRISP-DM” ซึ่งเกิดจากความร่วมมือระหว่างบริษัท DaimlerChrysler บริษัท SPSS และบริษัท NCR โดยมีโมเดลดังรูปที่ 1



รูปที่ 1. ขั้นตอนการทำงาน Model ของ CRISP-DM [1]

จากรูปที่ 1 แสดงขั้นตอนการทำงานของ CRISP-DM ดังนี้

2.2.1. Business Understanding เป็นขั้นตอนแรกสุดในกระบวนการ CRISP-DM ขั้นตอนนี้เป็นการทำความเข้าใจ ระบุปัญหาหรือโอกาสเชิงธุรกิจ

2.2.2. Data Understanding ในการรวบรวมข้อมูลนั้นควรพิจารณาด้วยว่าเป็นข้อมูลที่ไดมาจากแหล่งข้อมูลที่ต้องการ น่าเชื่อถือ ข้อมูลที่ได้มีปริมาณมากพอหรือยัง และเป็นข้อมูลที่เหมาะสม มีรายละเอียดเพียงพอต่อการนำไปใช้ในการวิเคราะห์

2.2.3. Data Preparation ขั้นตอนการเตรียมข้อมูลเป็นขั้นตอนที่ใช้เวลานานที่สุดเนื่องจากโมเดลที่ได้จากการทำเหมืองข้อมูลจะให้ผลลัพธ์ที่ถูกต้องหรือไม่ขึ้นขึ้นอยู่กับคุณภาพของข้อมูลที่ใช้การเตรียมมี 3 ขั้นตอนย่อยคือ ทำการคัดเลือกข้อมูล การกลั่นกรองข้อมูล การแปลงรูปข้อมูล

2.2.4. Modeling ขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูลได้แก่การสร้างตัวแบบทำนาย (Prediction model) ในบางครั้งพบว่ามีกรนำเทคนิคเหมืองข้อมูลหลายเทคนิคมาใช้ในการวิเคราะห์ข้อมูล เพื่อให้ได้ผลลัพธ์ที่ดีที่สุด

2.2.5. Evaluation การประเมินหรือวัดประสิทธิภาพของโมเดลวิเคราะห์ข้อมูลในขั้นตอนก่อนหน้านั้นเป็นเพียงการวัดความน่าเชื่อถือของโมเดล

2.2.6. Deployment ผลลัพธ์หรือองค์ความรู้ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล

2.3. การจำแนกชนิดข้อมูล [2][3]

การจำแนกชนิดข้อมูล (Classification) เป็นเทคนิคสำหรับการเรียนรู้แบบจำเป็นต้องมีผู้สอน (Supervised Learning) ซึ่งลักษณะการทำงานจะมีการแบ่งชุดของข้อมูลออกเป็น 2 ชุด โดยชุดข้อมูลส่วนแรกเรียกว่า “ชุดข้อมูลสำหรับการฝึกสอน” (Training set) เป็นชุดข้อมูลที่เอาไว้สำหรับให้เครื่องจักรเรียนรู้และสร้างตัวแบบการจำแนกข้อมูลขึ้นมา และชุดข้อมูลส่วนที่สองเรียกว่า “ชุดข้อมูลทดสอบ” (Testing set) เป็นชุดข้อมูลที้นำไปทดสอบประสิทธิภาพของตัวแบบการจำแนกชนิดข้อมูลที่ถูกรวบรวมขึ้นมา โดยเทคนิคพื้นฐานที่นิยมทั่วไปคือเทคนิคต้นไม้ตัดสินใจ (Decision Tree) และใช้อัลกอริทึมสำหรับการสร้างต้นไม้ตัดสินใจด้วย C4.5 ซึ่งมีการใช้ค่า Information Gain (IG) สำหรับตัดสินใจเลือกแอตทริบิวต์ในการสร้างต้นไม้ตัดสินใจ ดังสมการที่ 1 และสมการที่ 2

$$Gain(A) = Info(D) - Info_A(D). \quad (1)$$

$$Info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j). \quad (2)$$

2.4. งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษางานวิจัยต่างๆที่เกี่ยวข้องมีดังนี้ กิตติศักดิ์ สุมาลย์ [4] นำเสนอการวิเคราะห์ข้อมูลจากแบบสำรวจสภาวะสุขภาพประชาชนโดยใช้ข้อมูลของประชาชนในเขตเทศบาลเมืองจังหวัดบุรีรัมย์ แบ่งการศึกษาเป็น ประเด็นคือ การศึกษา 2 ความสัมพันธ์ของการเป็นโรคต่างๆ โดยใช้เทคนิค Association Rule และศึกษาการคัดกรองสุขภาพเบื้องต้นของประชาชนโดยแบ่งออกเป็น กลุ่มปกติ กลุ่มเสี่ยง กลุ่มป่วย ด้วยเทคนิค Classification นิเวศ จิระวิชิตชัย [5] นำเสนอการค้นหาเทคนิคเหมืองข้อมูลเพื่อสร้างโมเดลการวิเคราะห์โรคอัตโนมัติ มีวัตถุประสงค์เพื่อค้นหาเทคนิคด้านเหมืองข้อมูล เพื่อสร้างโมเดลการวิเคราะห์โรคอัตโนมัติทดสอบประสิทธิภาพในการจำแนก (Classification) มาลีวรรณ บุญพลอย [7] นำเสนอการศึกษาเรื่องเหมืองข้อมูลและการค้นหาองค์ความรู้ สำหรับโรคกล้ามเนื้ออ่อนแรง นำเสนอวิธีการใช้ขั้นตอนวิธีต้นไม้ตัดสินใจสองชั้นเพื่อจำแนกข้อมูลโรคกล้ามเนื้ออ่อนแรงเรื้อรัง เรวดี ศักดิ์ดุยธรรม [8] นำเสนอการประยุกต์ใช้เทคนิคดาต้าไมนิงในการวิเคราะห์ปัจจัยที่มีผลต่อความสำเร็จในการรักษาโรคนี้ในลักษณะต่างๆของคณะแพทยศาสตร์ศิริพยาบาล มหาวิทยาลัยกรุงเทพมหานคร เพื่อนำไปใช้ประกอบการตัดสินใจของแพทย์และบุคคลทั่วไปในการเลือกแนวทางในการรักษาโรคนี้ในลักษณะและอัครช บรรจงศิลป์ [10] ได้นำเสนอการศึกษาการหารูปแบบ

การพยากรณ์ที่เหมาะสมกับโรงงาน อุตสาหกรรมแต่ละประเภท เพื่อนำไปใช้ในการวางแผนการใช้พลังงานไฟฟ้าของผู้ใช้ไฟรายใหญ่

3. ขั้นตอนการดำเนินงาน

การวิจัยนี้ได้ดำเนินงานโดยประยุกต์ตามแนวทางในการทำเหมืองข้อมูลที่เรียกว่ากระบวนการมาตรฐานอุตสาหกรรม หรือ CRIPS-DM (Cross-Industry Standard Process for Data Mining) ในการสร้างโมเดลพยากรณ์การเกิดโรคเบาหวาน ผู้วิจัยได้ทำการกำหนดขั้นตอนและวิธีการดำเนินการวิจัยออกเป็น ขั้นตอน ดังนี้ 6

3.1 การวิเคราะห์ปัญหา (Business Understanding)

โรคเบาหวานเป็นโรคติดต่อเรื้อรังที่เป็นปัญหาสาธารณสุขที่สำคัญ และถือว่าเป็น “ภัยเงียบ” เพราะเป็นโรคที่ไม่ปรากฏอาการ และเป็นสาเหตุของโรคแทรกซ้อนในอวัยวะสำคัญหลายระบบของร่างกาย ดังนั้นจึงต้องมีการคัดกรองโรคเบาหวานเพื่อสร้างความตระหนักในการดูแลสุขภาพและได้รับการปรับเปลี่ยนพฤติกรรมสุขภาพเพื่อลดปัจจัยเสี่ยงต่อการเกิดโรคเรื้อรัง

3.2 การทำความเข้าใจกับข้อมูล (Data Understanding)

เพื่อหาปัจจัยที่มีผลต่อการพัฒนาตัวแบบประเมินความเสี่ยงโรคเบาหวาน โดยรวบรวมข้อมูลจากแหล่งข้อมูลที่ต้องการ นำเชื่อถือและข้อมูลที่เหมาะสม โดยได้รับความอนุเคราะห์จากโรงพยาบาลส่งเสริมสุขภาพตำบลตะคุ โดยได้ให้ศึกษาข้อมูลจริงของประชาชนในตำบลตะคุที่ได้รับการคัดกรองความเสี่ยงโรคเบาหวาน โดยเก็บข้อมูลจำนวน 1,000 เรคคอร์ด ประกอบด้วยแอตทริบิวต์ HN, เพศ, อายุ, น้ำหนัก, ส่วนสูง, BMI, รอบเอว, ความดันโลหิต เป็นต้น

3.3 การเตรียมข้อมูล (Data Preparation)

เป็นขั้นตอนการจัดการข้อมูลให้มีคุณภาพก่อนขั้นตอนการสร้างโมเดลซึ่งในงานวิจัยขั้นนี้ได้มีการทำความสะอาดข้อมูล (Data Cleaning) เพื่อแก้ปัญหาเรื่องข้อมูลที่ขาดหายไป (Missing Value) และทำการตัดแอตทริบิวต์หรือปัจจัยที่ไม่มีผลต่อการเกิดโรคเบาหวานออกไปเช่น แอตทริบิวต์ชื่อและแอตทริบิวต์ที่อยู่ เป็นต้น การแปลงค่าข้อมูลจากข้อมูลตัวเลข (ข้อมูลแบบต่อเนื่อง) ให้กลายเป็นข้อมูลแบบช่วงค่า (ข้อมูลแบบไม่ต่อเนื่อง) เพื่อให้เหมาะสมกับเทคนิคต้นไม้ตัดสินใจ และมีการกำหนดคลาสข้อมูลออกเป็น 3 คลาสคือ กลุ่มปกติ กลุ่มเสี่ยงและกลุ่มเสี่ยงมาก นอกจากนี้จำเป็นต้องมีการแปลงรูปแบบของโครงสร้างข้อมูลให้อยู่ในรูปแบบของ ARFF เพื่อให้เหมาะสมกับการนำไปสร้างโมเดลบนโปรแกรม WEKA

3.4 การสร้างโมเดล (Modeling)

งานวิจัยนี้ได้เลือกใช้ซอฟต์แวร์ WEKA สำหรับการสร้างโมเดลต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 โดยนำข้อมูลที่แปลงจากฐานข้อมูลใน Excel เป็นรูปแบบ ARFF มาสร้างตัวแบบทำนายโดยใช้เทคนิคต้นไม้ตัดสินใจ โดยข้อมูลได้จากการจัดเตรียมประกอบด้วยแอทริบิวต์ ดังรูปที่ 2

ลำดับ	ชื่อข้อมูล	รายละเอียด
1	Sex	เพศ
2	Ege	อายุ
3	BMI	ค่าดัชนีมวลกาย
4	Waist	ขนาดรอบเอว
5	Bps	ค่าความดันโลหิต
6	Drink	การดื่มสุรา
7	Smoke	การสูบบุหรี่
8	Exercise	การออกกำลังกาย
9	Eat1	อาหารรสหวาน
10	Eat2	อาหารประเภทของมัน
11	Eat3	อาหารรสเค็ม
12	Eat4	ไม่ชอบทุกข้อ
13	Fa1	ประวัติการเป็นเบาหวานของบิดา-มารดา
14	Fa2	ประวัติการเป็นโรคความดันโลหิตสูงของบิดา-มารดา
15	Fb1	ประวัติการเป็นเบาหวานของพี่น้องสายตรง
16	Fb2	ประวัติการเป็นโรคความดันโลหิตสูงของพี่น้องสายตรง
17	Result	ผลการประเมินโรคเบาหวาน

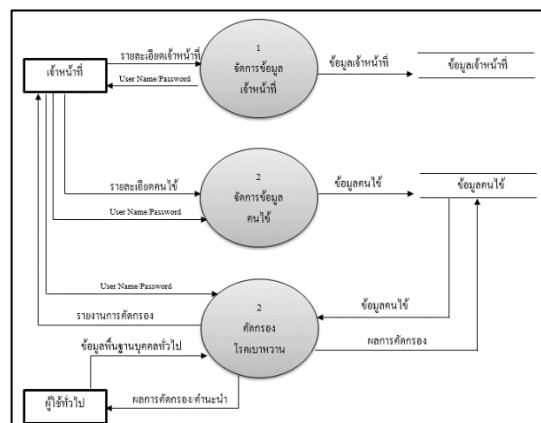
รูปที่ 2. ชุดข้อมูลสมมุติฐานก่อนสร้างตัวแบบพยากรณ์

3.5 การวัดประสิทธิภาพ (Evaluation)

งานวิจัยนี้ได้ทำการวัดประสิทธิภาพของโมเดลด้วยวิธีการแบ่งแยกข้อมูลตามร้อยละ (Percentage Split) แบบ 80:20 โดยใช้ข้อมูล 800 ชุดข้อมูล (ร้อยละ 80) สำหรับการฝึกสอนเพื่อสร้างตัวแบบการพยากรณ์การเกิดโรคเบาหวาน และใช้ข้อมูล 200 ชุดข้อมูล (ร้อยละ 20) สำหรับการทดสอบเพื่อประเมินประสิทธิภาพของตัวแบบ

3.6 การพัฒนาระบบการคัดกรองกลุ่มเสี่ยงโรคเบาหวาน

ในขั้นตอนนี้เป็นขั้นตอนการนำตัวแบบการพยากรณ์การเกิดโรคเบาหวานมาสร้างเป็นระบบสารสนเทศการคัดกรองกลุ่มเสี่ยงโรคเบาหวานในรูปแบบออนไลน์ด้วยภาษา PHP โดยสามารถแสดงโครงสร้างการทำงานด้วยแผนภาพกระแสข้อมูล (Data Flow Diagram) ดังรูปที่ 3



รูปที่ 3. แผนภาพกระแสข้อมูล ระดับ 1 ระบบการคัดกรองกลุ่มเสี่ยงโรคเบาหวาน

4. ผลการดำเนินงาน

ผลการดำเนินงานสามารถแบ่งออกได้เป็น 3 ส่วนประกอบด้วย ส่วนผลการประเมินประสิทธิภาพของตัวแบบการพยากรณ์การเกิดโรคเบาหวาน ส่วนผลการทำงานของระบบการคัดกรองกลุ่มเสี่ยงโรคเบาหวาน และส่วนผลการประเมินความพึงพอใจการทำงานของพนักงาน

4.1 ส่วนผลการประเมินประสิทธิภาพของตัวแบบการพยากรณ์การเกิดโรคเบาหวาน

งานวิจัยนี้ได้ทำการทดสอบประสิทธิภาพของตัวแบบการพยากรณ์การเกิดโรคเบาหวานด้วยวิธีการแบ่งแยกข้อมูลตามร้อยละซึ่งมีการทดสอบประสิทธิภาพในรูปแบบ 80:20 และ 70:30 การทดลองจะทำการแบ่งแยกข้อมูลแบบเลือกชุดข้อมูลเองจำนวน 3 ชุดข้อมูล โดยได้ค่าความถูกต้อง ดังตารางที่ 1

ตารางที่ 1. ประสิทธิภาพความถูกต้องของตัวแบบพยากรณ์

ผลเปรียบเทียบ	Split 80:20	Split 70:30
ตัวแบบที่ 1	69%	66.67%
ตัวแบบที่ 2	72.5%	70%
ตัวแบบที่ 3	74.5%	73%

จากตารางที่ 1 แสดงผลการประเมินประสิทธิภาพของตัวแบบพยากรณ์การเกิดโรคเบาหวานซึ่งผลการทดลองแสดงให้เห็นว่าค่าความถูกต้องในการพยากรณ์การเกิดโรคเบาหวานของตัวแบบที่ 3 มีค่าความถูกต้องสูงที่สุดในรูปแบบการแบ่งแยกชุด

ข้อมูลแบบ 80:20 ซึ่งมีความใกล้เคียงกับค่าความถูกต้องใน
รูปแบบการแบ่งแยกชุดข้อมูลแบบ 70:30

4.2 ส่วนผลการทำงานของระบบการคัดกรองกลุ่มเสี่ยงโรคเบาหวาน

การทำงานของระบบการคัดกรองกลุ่มเสี่ยงโรคเบาหวานสามารถ
แบ่งรูปแบบการคัดกรองกลุ่มเสี่ยงโรคเบาหวานแบบสำหรับ
บุคคลทั่วไปและสำหรับเจ้าหน้าที่

4.2.1 ระบบการคัดกรองกลุ่มเสี่ยงโรคเบาหวานสำหรับบุคคลทั่วไป การทำงานจะสามารถให้บุคคลทั่วไปเข้าไปกรอกข้อมูลพื้นฐานส่วนบุคคลบนหน้าฟอร์มการคัดกรองเพื่อแสดงผลการคัดกรองออกมา แสดงได้ดังรูปที่ 4 และรูปที่ 5



รูปที่ 4. หน้าฟอร์มการกรอกข้อมูลการคัดกรองโรคเบาหวาน



รูปที่ 5. หน้าแสดงผลการกรอกข้อมูลการคัดกรองโรคเบาหวาน

4.2.2 ระบบการคัดกรองกลุ่มเสี่ยงโรคเบาหวานสำหรับเจ้าหน้าที่
การทำงานจะให้เจ้าหน้าที่โรงพยาบาลสามารถนำข้อมูลของ

คนใช้เข้าสู่ระบบการคัดกรองเพื่อแสดงผลการคัดกรองออกมา
แสดงได้ดังรูปที่ 6 และรูปที่ 7

[illegible]

รูปที่ 6. หน้าฟอร์มการคัดกรองกลุ่มผู้เสี่ยงโดยเจ้าหน้าที่

4.3 ส่วนผลการประเมินความพึงพอใจของผู้ใช้

การทดสอบระบบเพื่อหาความพึงพอใจในการใช้งานระบบ การประเมินความพอใจของระบบใช้การประเมินความพึงพอใจ ด้วยมาตราอันดับ 5 อันดับ และแปลความหมายของค่าเฉลี่ยเป็น แบบ ช่วงคะแนน ดังตารางที่ 52 โดยมีการทดสอบกับผู้ใช้งานจำนวน 3 คน ประกอบด้วยเจ้าหน้าที่โรงพยาบาล

ตารางที่ 2. เกณฑ์การแปลผลการประเมินความพึงพอใจ

เกณฑ์ ค่าเฉลี่ย	การแปลความหมายของระดับ คะแนน
4.51-5.00	เกณฑ์ดีมาก
3.51-4.50	เกณฑ์ดี
2.51-3.50	เกณฑ์พอใช้
1.51-2.50	เกณฑ์ปรับปรุง
1.00-1.50	เกณฑ์ไม่เหมาะสม

ซึ่งสถิติที่ใช้ในการวิเคราะห์ข้อมูลความพึงพอใจต่อระบบมี 2 สมการ คือ ค่าเฉลี่ยเลขคณิต (Mean) และส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (3)$$

$$S.D. = \sqrt{\frac{n \sum f x_i^2 - (\bar{x})^2}{n(n-1)}} \quad (4)$$

ผลการประเมินความพึงพอใจของผู้ใช้งาน มีความพึงพอใจต่อระบบสารสนเทศการคัดกรองกลุ่มเสี่ยงโรคเบาหวานในภาพรวมของระบบอยู่ในเกณฑ์พอใจเท่ากับ 3.06 ดังตารางที่ 3

ตารางที่ 3. ผลการประเมินความพึงพอใจ

รายการประเมิน	□	S.D.	ระดับความพึงพอใจ
1. ใช้งานง่าย เรียนรู้ได้เร็ว สะดวกต่อการใช้งาน	3.33	0.33	พอใช้
2. แบ่งการใช้งานเป็นหมวดหมู่	3	0	พอใช้
3. ความถูกต้องและสมบูรณ์ของระบบ	3	0	พอใช้
4. ความรวดเร็วในการใช้บริการ	3.66	0.33	ดี
5. ความสะดวกรวดเร็ว	3	0	พอใช้
6. ความทันสมัย น่าสนใจ เป็นปัจจุบัน	3	0.33	พอใช้
7. สามารถนำไปใช้ในการบริหาร และการตัดสินใจ	3	1	พอใช้
8. สะดวกในการปรับปรุงแก้ไข ในระยะยาว ตามความก้าวหน้าของเทคโนโลยี	3	0	พอใช้
9. การให้คำแนะนำ และแก้ปัญหา	3	0	พอใช้
10. ระบบมีความพร้อมในการให้บริการ	3	0	พอใช้

5. สรุปผลและข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างโมเดลที่มีความเหมาะสมและถูกต้อง สำหรับพัฒนาระบบการคัดกรองโรคเบาหวาน โดยใช้เทคนิคต้นไม้ตัดสินใจ อัลกอริทึม C4.5 (J48) ซึ่งได้ทำการวัดประสิทธิภาพของโมเดลด้วยวิธี percentage split ได้ใช้ข้อมูลที่แบ่ง 80% ใช้สร้างโมเดลหรือข้อมูลที่จะเป็นข้อมูลเรียนรู้ และเพื่อเป็นการวัดประสิทธิภาพของโมเดล ข้อมูล สำหรับ 20% (การเรียนรู้ Training Data) และ ชุดข้อมูลทดสอบ (Testing Data) เป็นร้อยละ 1 จากชุดข้อมูลทั้งหมด 20 และ 80,000 ชุดข้อมูลคือ 800 และ 200 ของชุดข้อมูลตามลำดับ ได้ค่า Correctly Classified Instances 74.5% ในอนาคตผู้วิจัยได้วางแผนในการปรับปรุงการทำงานของระบบโดยจะประยุกต์ใช้เทคนิคการเรียนรู้แบบอื่นเข้ามาเพื่อเปรียบเทียบประสิทธิภาพที่ดีที่สุด

6. เอกสารอ้างอิง

- [1] Chapman Pete, Clinton Julian, Kerber Randy, eds. CRISP-DM 1.0 Step-by-step data mining guide. SPSS Inc., USA. 2000.
- [2] Jaiwei Han and Micheline Kamber. Data Mining Concepts and Techniques, Second Edition. University of Illinois at Urbana-Champaign. 2006.
- [3] S. Prabhu and N. Venkatesan. Data mining and Warehousing. New Age International Publicshers. 2007.
- [4] กิตติศักดิ์ สุमार. (2555). การคัดกรองสุขภาพเบื้องต้นโดยใช้เทคนิคเหมืองข้อมูล.
- [5] นิเวศ จิระวิชัย. (2553). การค้นหาเทคนิคเหมืองข้อมูลเพื่อสร้างโมเดลการวิเคราะห์โรคอ้วน
- [6] ณัฐพงษ์ วาริประเสริฐ. (2552). ปัญญาประดิษฐ์ (Artificial Intelligence). 1. กรุงเทพฯ : เคทีพี คอมพ์ แอนด์ คอนซัลท์.
- [7] มาลีวรรณ บุญพลอย. (2551). เหมืองข้อมูลและการค้นหาองค์ความรู้สำหรับโรคกล้ามเนื้ออ่อนแรง ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยเกษตรศาสตร์.
- [8] เรวดี ศักดิ์กุลธรรม. (2553). วิเคราะห์ปัจจัยที่มีผลต่อความสำเร็จในการรักษาโรคนี้ในแบบต่างๆ.
- [9] สายชล สันสมบูรณ์ทอง. (2558). การทำเหมืองข้อมูล Data Mining. 1 กรุงเทพฯ : บริษัท จามจุรีโปรดักส์ จำกัด 26 ซอยพระราม2 ที่ 83 ถนนพระราม2 ซอย ส้อมดำ เขตบางขุนเทียน.
- [10] อัคริช บรรจงศิลป์. (2550). การพยากรณ์การใช้ไฟฟ้าในอุตสาหกรรมขนาดใหญ่