

การวิเคราะห์ความเสี่ยงของคนเป็นโรคข้อเข่าเสื่อมด้วยเทคนิคเหมืองข้อมูล

Risk Analysis of Osteoarthritis Using Data Mining

เบญญาภา ศรีสว่าง, สุภาวดี ทองคำและ สุพาพร บรรดาศักดิ์

Benyapha Seesawang, Supawadee Thongkum, Supaporn Bundasak

jumbundasak@hotmail.com, benyapha.se@ku.th, supawadee.miew@hotmail.com

ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ ศรีราชา มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา

199 ถนนสุขุมวิท ตำบลทุ่งสุขลา อำเภอศรีราชา ชลบุรี 20230

บทคัดย่อ

ในส่วนงานวิจัยเชิงโปรแกรม เราได้เลือกเรื่องการวิเคราะห์โรคข้อเข่าเสื่อมมาเป็นส่วนหนึ่งของการพัฒนาระบบในการวิเคราะห์ความเสี่ยงของโรคซึ่งก็คือระบบการวิเคราะห์หาความเสี่ยงของโรคข้อเข่าเสื่อมโดยจะมีความสำคัญต่อผู้ใช้งาน และง่ายต่อการใช้งาน สะดวกต่อการทำความเข้าใจและจะทำให้คนหันมาใส่ใจดูแลตนเองมากขึ้น เพื่อเสี่ยงต่อการที่จะเกิดความเสี่ยงของโรคข้อเข่าเสื่อม

ผลการวัดประสิทธิภาพของระบบจากการหาค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยสามารถสรุปได้ว่าระบบการวิเคราะห์ความเสี่ยงในการเกิดโรคข้อเข่าเสื่อมใกล้เคียงกับความจริงโดยใช้โมเดลNaïve Bayes ทั้ง 2กลุ่มมีประสิทธิภาพค่าความถูกต้องมีค่าเท่ากับ 92.1466 %, 99.7382 % ตามลำดับ จะเห็นได้ว่าโมเดลNaïve Bayes มีประสิทธิภาพเหมาะสมกับข้อมูลที่ใช้ในการวิเคราะห์ความเสี่ยงต่อโรคข้อเข่าเสื่อม และระบบนี้จะส่งผลดีและเป็นประโยชน์ให้กับผู้ใช้ทั่วไปและบุคคลที่มีความเสี่ยงที่จะเป็นโรคข้อเข่าเสื่อมด้วย

คำสำคัญ โรคข้อเข่าเสื่อม; เหมืองข้อมูล; การจำแนกข้อมูล; Naïve byes; การพยากรณ์โรค**

Abstract

In research programs. We have chosen the subject analysis of osteoarthritis as part of developing a system to analyze the risks of the disease, that is, systems analysis, the risk of osteoarthritis by will are important to users and easy to use. Easy to understand and to make people turn to care themselves more to avoid the, there is a risk of osteoarthritis. Measurement of system performance from the average of the absolute tolerance can conclude that the risk analysis system in osteoarthritis, closer to truth by Naïve Bayes model. Both the powerful group 2, the correct value is equal to 92.1466%, 99.7382%, respectively, it can be seen that Naïve Bayes model. Effective, appropriate to the data used in the analysis of the risk of osteoarthritis, and this system will have a positive impact and benefits to users in General, and the person is at risk, as with osteoarthritis.

Keywords: Data Mining Technique, Osteoarthritis

บทนำ

ในปัจจุบันการใช้ชีวิตของคนในสังคมไม่ว่าจะเป็น การทำงานการใช้ชีวิตประจำวันล้วนมีความเสี่ยงต่อการเกิดโรคร้ายไข้เจ็บมากมายโดยเฉพาะโรคข้อเข่าเสื่อมที่ส่วน ใหญ่แล้วคนมักจะละเลยการดูแลตัวเองด้านการใช้ข้อเข่าโรค นี้จึงมักจะเกิดขึ้นกับวัยผู้สูงอายุ ซึ่งพฤติกรรมต่างๆใน ชีวิตประจำวันอาจทำให้เกิดโรคข้อเข่าเสื่อม อาทิเช่นการนั่ง ยองหรือพับเพียบเป็นเวลานาน การทำงานที่อยู่ในพื้นชั้นและ การเดินขึ้นลงจากบันไดบ่อยครั้งก็อาจทำให้เกิดภาวะความ เสี่ยงที่จะเป็นโรคข้อเข่าเสื่อมหรือข้อเข่าอักเสบได้ด้วยเช่นกัน

เราจึงได้เลือกโรคข้อเข่าเสื่อมมาเป็นส่วนหนึ่งของการ พัฒนาระบบในการวิเคราะห์ความเสี่ยงของโรคซึ่งก็คือ ระบบการวิเคราะห์หาความเสี่ยงของโรคข้อเข่าเสื่อมโดยจะมี ความสำคัญต่อผู้ใช้งานและง่ายต่อการใช้งานสะดวกต่อการ ทำความเข้าใจและจะทำให้คนหันมาใส่ใจดูแลตนเองมากขึ้น เพื่อเลี่ยงต่อการที่จะเกิดความเสี่ยงของโรคข้อเข่าเสื่อมซึ่ง ระบบนี้จะมีการทำงานในลักษณะการสอบถามอาการแรกเริ่ม ที่จะส่งผลต่อการเกิดความเสี่ยงโรคข้อเข่าเสื่อมอายุ เพศ และ ปัจจัยต่างๆที่เกี่ยวข้องและมีการอธิบายการดูแลตัวเอง ความรู้ความเข้าใจและปัจจัยต่างๆในโรคข้อเข่าเสื่อมและ ระบบนี้จะส่งผลและเป็นประโยชน์ให้กับผู้ใช้ทั่วไปและบุคคล ที่มีความเสี่ยงที่จะเป็นโรคข้อเข่าเสื่อมด้วย

2. วัตถุประสงค์

- 2.1. เพื่อพัฒนาการวิเคราะห์หาความเสี่ยงของโรคข้อเข่า เสื่อมว่าใช้แบบจำลองใดในการวัดหาระดับความเจ็บปวดของ ผู้มีความเสี่ยง
- 2.2. เพื่อเปรียบเทียบอัลกอริทึมว่ามีประสิทธิภาพที่เหมาะสม ในการวิเคราะห์ความเสี่ยงของโรคข้อเข่าเสื่อม

3. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

3.1. ทฤษฎีที่เกี่ยวข้อง

งานวิจัยของระบบวิเคราะห์ความเสี่ยงการเกิดโรคข้อเข่า เสื่อมนี้ได้นำ 4 อัลกอริทึม จากเทคนิคเหมืองข้อมูลคือ Decision tree (J4 8) ,Naïve bytes, Sequential Minimal Optimization(SMO) และ Neural network เข้ามาใช้ในการ วิเคราะห์เพื่อหาปัจจัยหาความสัมพันธ์ระหว่างข้อมูล พฤติกรรมและระดับความเจ็บปวดข้อเข่าของคนที่เป็นโรคข้อ เข่าเสื่อม และผลการวิเคราะห์ที่ได้ซึ่งก็คือ อัลกอริทึม Naïve bytes โดยนำไปเป็นส่วนประกอบส่วนหนึ่งของ Web Application ช่วยในการวิเคราะห์ความเสี่ยงในการเกิดโรคข้อ เข่าเสื่อม

3.1.1 ข้อเข่าเสื่อม

ข้อเข่าเสื่อม หมายถึง โรคที่เกิดจากความเสื่อมของ กระดูกอ่อนผิวข้อ ทั้งทางด้านรูปร่าง โครงสร้าง การทำงาน ของกระดูกข้อต่อและกระดูกบริเวณใกล้เคียง การเปลี่ยนแปลง ที่เกิดขึ้นไม่สามารถกลับสู่สภาพเดิมและอาจมีความเสื่อม รุนแรงขึ้นตามลำดับจากการศึกษาเชิงวิเคราะห์ความสัมพันธ์ มีวัตถุประสงค์เพื่อทำนายความเสี่ยงการเกิดโรคข้อเข่าเสื่อม ของผู้ป่วยสรุปผลการวิจัยได้ว่าการวัดระดับความรุนแรง ของข้อเข่าเสื่อมได้ทั้งหมด3 ระยะคือ

ระยะที่ 1 เป็นระยะที่มีความรุนแรงของโรคน้อยเริ่ม จากมีอาการปวดเมื่อยเล็กน้อยๆหลังจากการทำงานหนัก เดินไหวและอาการจะค่อยๆเพิ่มมากขึ้นแต่พอก้าวเดินไป นานๆสัก 10-15 ก้าวอาการก็จะเริ่มหายปวดหรือเรียกได้ว่า เป็นครั้งคราวหรือนานๆครั้ง

ระยะที่ 2 เป็นระยะที่มีความรุนแรงของโรคข้อเข่า เสื่อมปานกลางอาการปวดจะมากขึ้นและบ่อยขึ้นจนทำให้ขึ้น ลงบันไดลำบากนั่งยองๆหรือนั่งพับเพียบนานไม่ได้เมื่อลุก ต้องมีคนพยุงอาจกินเวลา5-15ปี

ระยะที่ 3 เป็นระยะที่มีความรุนแรงของโรคข้อเข่า เสื่อมมากโดยมีอาการปวดเข่าอยู่ตลอดเวลาไม่ว่าจะพับข้อ เข่าใช้ข้อหรือเรียกได้ว่าทุกอย่างก้าวที่เคลื่อนไหวต้องมีคนช่วย พยุงตลอดเวลาและไม่สามารถช่วยเหลือตัวเองได้

3.1.2 การทำเหมืองข้อมูล

Data mining เป็นกระบวนการของการค้นพบองค์ความรู้ที่ น่าสนใจจากจำนวนข้อมูลที่มีมากมายที่ถูกเก็บไว้ใน ฐานข้อมูล Data Warehouses หรือที่เก็บข้อมูลอื่นๆช่วยให้

สามารถทำนายแนวโน้มของข้อมูลใหม่ที่จะเกิดขึ้นในอนาคต
ได้การทำเหมืองข้อมูลแบ่งได้ 4 ขั้นตอนหลักดังนี้

ขั้นตอนที่ 1 เตรียมข้อมูล (Data Preparation) ถ้าข้อมูล
ไม่อยู่ในรูปแบบที่ถูกต้องหรือเหมาะสมจะต้องมีการปรับ
ข้อมูลให้อยู่ในรูปแบบเดียวกันและจัดข้อมูลที่ไม่ถูกต้อง
ออกไป

ขั้นตอนที่ 2 ลดขนาดของข้อมูล (Data Reduction) :
เนื่องจากถ้าข้อมูลมีปริมาณมากเกินไปการค้นหามODEL หรือ
แพทเทิร์นจากกลุ่มข้อมูลขนาดใหญ่ต้องใช้เวลามากซึ่งถาลด
จำนวนข้อมูลลงด้วยสัดส่วนที่ถูกต้องโมเดลที่ได้ยังคงเป็น
เช่นเดิมจะช่วยเพิ่มประสิทธิภาพมากขึ้น

ขั้นตอนที่ 3 ค้นหาโมเดล (หรือความสัมพันธ์) จาก
ข้อมูล (Data Modeling/Discovery) กระบวนการค้นหาโมเดล
หรือความสัมพันธ์จะเริ่มจากข้อมูลเริ่มต้นจำนวนไม่มากนัก
จากนั้นผลที่ได้จากกระบวนการค้นหา (Learning
Process/Method) ไปยืนยันกับข้อมูลทดสอบถ้าผลที่ได้ยังไม่
น่าพอใจอาจจะต้องปรับค่าพารามิเตอร์บางตัวของ Learning
Method และเริ่มกระบวนการค้นหาใหม่กับข้อมูลจำนวนมาก
ขึ้นจนกว่าผลที่ได้มีความถูกต้องอยู่ในระดับที่ยอมรับได้จึงจะ
จบกระบวนการค้นหา

ขั้นตอนที่ 4 ตรวจสอบและวิเคราะห์ผล (Solution
Analyses) : โมเดลหรือความสัมพันธ์ที่หาได้ในขั้นตอนที่
3 จะต้องถูกมาทดสอบอัตราความผิดพลาดและวิเคราะห์
ความซับซ้อนของรูปแบบโมเดลถ้าอัตราความผิดพลาดยังสูง
เกินไปอาจจะต้องย้อนกลับไปขั้นตอนที่ 3 อีกครั้งเพื่อ
ปรับปรุงโมเดลให้ถูกต้องยิ่งขึ้น

3.1.3 เทคนิค Naive Bayes

Naive Bayes คือโมเดลการคัดแยกประเภทข้อมูลที่ใช้หลัก
ความน่าจะเป็นซึ่งอยู่บนพื้นฐานของ Bayes' Theorem และ
สมมติฐาน ที่ให้การเกิดของเหตุการณ์ต่างๆเป็นอิสระต่อกัน
(independence)

กำหนดให้ P(H) ความน่าจะเป็นที่จะเกิดเหตุการณ์ H
และ P(H|E) คือความน่าจะเป็นที่จะเกิดเหตุการณ์ H เมื่อเกิด
เหตุการณ์ E จากตัวแปรที่กำหนดและแนวคิดของ Bayes'
Theorem นั้นเราสามารถทำนายเหตุการณ์ที่พิจารณาได้จาก
การเกิดของเหตุการณ์ต่างๆ ได้ดังสมการ

$$P(H|E) = [P(E|H) \times P(H)] / P(E) \quad (1)$$

3.1.4 Decision tree(J48)

ต้นไม้ตัดสินใจหรือเรียกอีกชื่อว่า Classification tree เป็น
เทคนิคแบบ Supervised Learning ที่มีวิธีเรียนรู้ที่ไม่ซับซ้อน
มากนัก โดยมีโครงสร้างเป็นต้นไม้ แยกแขนงไปตามเงื่อนไข
หรือข้อมูลที่ได้คาดคะเนไว้ว่าจะเกิดขึ้น โครงสร้างของ
Decision tree จะประกอบด้วย ใบ (Leaf) เป็นส่วนของ
ผลลัพธ์ที่ได้, กิ่ง (Branch) เป็นตัวเชื่อมระหว่างโหนด และ
ราก (Root) เป็นจุดเริ่มต้นของเหตุการณ์

การเรียนรู้ด้วย Decision tree เป็นการเรียนรู้จากการ
คาดคะเนเหตุการณ์ต่างๆเกิดจากการคำนึงถึงปัจจัยเป็น
เงื่อนไขช่วยในการตัดสินใจว่าเมื่อมีเหตุการณ์ใดๆ ขึ้น จะ
แสดงผลลัพธ์ออกมาอย่างไรต้องการต้นไม้ให้มีขนาดเล็ก,การ
คาดคะเนจะถูกนำเสนอด้วยรูปแบบกฎการตัดสินใจ
(if/then/else)

3.1.5 Sequential Minimal Optimization(SMO)

เป็นอัลกอริทึมสำหรับการแก้สมการกำลังสองการเขียน
โปรแกรม quadratic programming (QP) ซึ่งเป็นปัญหาที่
เกิดขึ้นในระหว่างการฝึกของการสนับสนุนเครื่องเวกเตอร์
(support vector machines) และเป็นที่นิยมเป็นอย่างมาก

3.1.6 Neural network

คือโมเดลทางคณิตศาสตร์สำหรับประมวลผลสารสนเทศด้วย
การคำนวณ แบบคอนเนชันนิสต์ (connectionist) เพื่อ
จำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ ด้วย
วัตถุประสงค์ที่จะสร้างเครื่องมือซึ่งมีความสามารถในการ
เรียนรู้การจดจำแบบรูป (Pattern Recognition) และการ
อุปมา ความรู้ เช่นเดียวกับความสามารถที่มีในสมองมนุษย์
แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาข่ายงาน
ไฟฟ้าชีวภาพ (bioelectric network) ในสมอง

3.1.7 Weka

โปรแกรมWeka เป็นซอฟต์แวร์ประเภทฟรีแวร์ซึ่งอยู่ภายใต้
การควบคุมของ GPL License ถูกพัฒนาขึ้นบนพื้นฐานของ
ภาษาจาวาโดยเน้นใช้งานทางด้านการเรียนรู้ด้วยเครื่อง
(Machine Learning) และการทำเหมืองข้อมูล (Data Mining)

มีโมดูลย่อยสำหรับใช้ในการจัดการข้อมูลและใช้ตัวแสดงผลผ่านทางหน้าจอในลักษณะ GUI (Graphic User Interface)

3.2. วิจัยที่เกี่ยวข้อง

จากการศึกษาพบว่ามิงงานวิจัยจำนวนไม่น้อย ได้พัฒนาขึ้นเพื่อนำมาแก้ปัญหามีผลการเกิดโรคหลายโรคผู้วิจัยจากการศึกษางานวิจัยที่มีความเกี่ยวข้องดังต่อไปนี้

กิตติพลวิแสงและคณะ[1] ได้วิจัยปัจจัยเสี่ยงการเกิดโรคเบาหวานแทนการตรวจเลือดเพื่อวิเคราะห์ปัจจัยเสี่ยงและลำดับการเกิดโรคเบาหวาน โดยเลือกใช้ตัวศึกษา Back-propagation Neural Networks, Radial Basis Function Network และตัวแบบ Naïve Bayes ผลจากการวินิจฉัยทางการแพทย์ได้ถูกนำมาอ้างอิงเพื่อวัดความถูกต้องของตัวแบบจากผลได้ตัวแบบ Back-propagation Network ได้ผลลัพธ์ที่ดีที่สุด เนื่องจาก ปัจจัยเสี่ยงและลำดับของปัจจัยเสี่ยงสามารถตัดสินใจได้จากตัวแบบที่ศึกษาและสามารถตรวจประเมินความเสี่ยงโรคเบาหวานโดยไม่ต้องอาศัยการตรวจเลือดได้รวมทั้งเป็นเครื่องมือในการประเมินตนเองอีกด้วย

รัตนวลัภกิตติสมัยและคณะ[2]จากงานวิจัยผู้เชี่ยวชาญสนใจที่จะศึกษาการพัฒนาการดูแลผู้สูงอายุที่มีปัญหาปวดข้อเข่าในชุมชน เพื่อให้ผู้สูงอายุสามารถที่จะจัดการกับอาการปวดได้อย่างเหมาะสมและมีความมั่นใจในการปฏิบัติกิจกรรมลดปวดแทนการใช้ยาแก้ปวดได้และเป็นการป้องกันความเสี่ยงที่จะเกิดขึ้นจากอาการปวดข้อและการใช้งานข้อผิดโดยนำทฤษฎีการสร้างเสริมสมรรถนะแห่งตนมาประยุกต์ใช้เพื่อให้ผู้สูงอายุมีความเชื่อมั่นว่าตนเองมีความสามารถในการปฏิบัติกิจกรรมการบริหารกล้ามเนื้อรอบเข่าเพื่อลดอาการปวดข้อเข่าแทนการใช้ยาได้

นพรัตน์พจน์ จิราภรณ์และคณะ[3]การวิจัยเชิงทดสอบโรคหลอดเลือดโป่งพองภาวะที่หลอดเลือดเปลี่ยนแปลงขยายขนาดอย่างถาวรโดยได้นำเรื่องนี้มาพัฒนาโดยเลือกใช้โมเดล Decision tree ,Naïve byes และ Neuron Network เพื่อที่จะหาความสัมพันธ์ระหว่างข้อมูลของผู้ป่วยประวัติการเจ็บป่วยในอดีตและโรคหลอดเลือดแดงโป่งพองจากการศึกษาของระบบที่พัฒนาขึ้นสามารถพยากรณ์โรคได้ถูกต้อง 87.57% และผลการสำรวจที่ประสิทธิภาพโดยรวมพบว่าความพึงพอใจเฉลี่ยอยู่ที่4.45 ที่ส่วนเบี่ยงเบนมาตรฐานที่ 0.56 ดังนั้นจึงสามารถสรุปได้ว่าระบบที่พัฒนาขึ้นนั้นมีประสิทธิภาพในระดับที่ดี

4.วิธีการดำเนินการวิจัย

4.1.ศึกษาปัญหาและวิเคราะห์ข้อมูล

จากการศึกษาและรวบรวมสรุปงานวิจัยที่เกี่ยวข้อง รวมถึงการศึกษาการทำงานของเทคนิคเหมืองข้อมูลเกี่ยวกับข้อมูลทางการแพทย์พิจารณาเทคนิคเหมืองข้อมูลในการวิจัยดังต่อไปนี้ Decision Tree , Naïve Bayes ,Neural networks และ Sequential Minimal Optimization(SMO)รวมทั้งวิธีการวิเคราะห์และประเมินผลศึกษาวิธีการใช้งานซอฟต์แวร์ที่ใช้ในการทำเหมืองข้อมูลและรวบรวมข้อมูลมาจากเก็บข้อมูลจากผู้ป่วยที่เป็นโรคข้อเข่าเสื่อมและคนที่ไม่เป็นโรคข้อเข่าเสื่อม โดยจะใช้วิธีการทำงานของเครื่องมือหรือซอฟต์แวร์ที่ใช้ในการทำเหมืองข้อมูลและทำการสร้างแบบสอบถามโดยอิงจากงานวิจัยที่เกี่ยวข้องกับโรคข้อเข่าเสื่อมโดยจะเลือกคำถามที่ส่งผลต่อการเกิดโรคข้อเข่าเสื่อมมากที่สุด และสร้างแบบจำลองการพยากรณ์โดยนำข้อมูลที่ได้จากแบบสอบถามเพื่อวิเคราะห์ความเสี่ยงของการเกิดโรคข้อเข่าเสื่อม และข้อมูลที่เก็บได้จากแบบสอบถามทั้งหมด 383 ชุดได้ชุดข้อมูลเพื่อทำการวิเคราะห์ 2 กรณี ชุดข้อมูลที่ 1 วิเคราะห์ความเสี่ยงหาปัจจัยกรณีคัดกรองผู้เป็นโรคจากข้อมูลคนที่เป็นโรคข้อเข่าเสื่อมจำนวน 193 คนและคนที่ไม่เป็นโรคข้อเข่าเสื่อมจำนวน 190คน ชุดที่2 วิเคราะห์กรณีที่ 2 สำหรับผู้ที่เสี่ยงเป็นโรคข้อเข่าเสื่อมรักษาตามระดับความเจ็บปวด3 ระดับ (1,2,3)จากข้อมูลคนที่เป็นโรคข้อเข่าเสื่อมจำนวน 193 คน แบ่งเป็นมีระดับความรุนแรงที่1 จำนวน 90,มีระดับความรุนแรงที่2 จำนวน 73คน และมีระดับความรุนแรงที่3จำนวน 30ชุด

4.2.การเตรียมข้อมูล

นำข้อมูลที่สุ่มมาจัดรูปแบบแฟ้มข้อมูลให้อยู่ในรูปแบบที่เหมาะสมเพื่อใช้กับซอฟต์แวร์ WEKA โดยดำเนินการป้อนข้อมูลลงเอ็กเซล และก่อนจะนำข้อมูลไปใช้วิเคราะห์และประมวลผลในซอฟต์แวร์ WEKA จะแปลงไฟล์ข้อมูลให้เป็น CSV data filesจากนั้นดำเนินการทำ Data cleaning โดยการลบ Record ที่มีค่าที่มีความผิดปกติหรือไม่จำเป็นต้องใช้ซึ่งเมื่อนำข้อมูลดังกล่าวไปวิเคราะห์อาจทำให้รูปแบบทางสถิติมีความคลาดเคลื่อนหรือผิดพลาดแล้วจัดการกับข้อมูลสูญหายโดยแทนที่ข้อมูลที่สูญหายด้วยค่าเฉลี่ยของเขตข้อมูลสำหรับข้อมูลที่เป็นตัวเลขและแทนด้วยฐานนิยมสำหรับข้อมูล

เชิงจำแนกสำหรับข้อมูลที่เกี่ยวข้องกับการแพทย์เช่นข้อมูลส่วนตัวข้อมูลการประเมินความเจ็บปวดข้อเข่าพฤติกรรมเสี่ยงต่อการเกิดข้อเข่าเสื่อมโดยแสดงการศึกษาในครั้งนี้ผู้ศึกษาเลือกใช้วิธีการตัดแถวที่ปรากฏ Missing value ซึ่งเป็นวิธีที่นิยมมากสำหรับเทคนิคการจำแนกข้อมูล โดยข้อมูลที่วิเคราะห์จะแบ่งตามตารางดังนี้

ตารางที่1 ข้อมูลทั่วไป (ข้อมูลส่วนตัว)

ลำดับ	ตัวแปร	ชื่อข้อมูล
1	Sex	เพศ
2	Age	อายุ
4	Weight	น้ำหนัก
5	Height	ส่วนสูง
6	BMI	ดัชนีมวลกาย
7	Job	อาชีพ
8	Exercise	การออกกำลังกาย
9	Smoke	การสูบบุหรี่

ตารางที่2 ข้อมูลพฤติกรรมที่เสี่ยงต่อโรค

ลำดับ	ตัวแปร	ชื่อข้อมูล
1	Walk	เดิน
2	Walk up stair	ขณะขึ้นบันได
4	Walk down stair	ขณะลงบันได
5	Sleeping	ขณะนอน
6	Sit up	ขณะลุกนั่ง
7	Stand	ขณะลุกจากที่นอน
8	Wake up	ขณะยืน
9	Change stance	ขณะเปลี่ยนท่านั่งเดิม
10	Bent down	ขณะก้มตัว
11	Up down car	ขณะขึ้น-ลงรถ
12	Wear shoesand socks	ขณะใส่รองเท้าถุงเท้า
13	Takeoff shoes and socks	ขณะถอดรองเท้าถุงเท้า
14	Get out of bed	ขณะลุกจากเตียง
15	Sitting	ขณะนั่ง
16	Sit toilet	ขณะลุกนั่งจากชักโครก
17	Household work	ขณะทำงานบ้าน

ตารางที่3 พฤติกรรมการเคลื่อนที่และการทานอาหารที่เสี่ยงการเกิดโรค

ลำดับ	ตัวแปร	ชื่อข้อมูล
1	FatFood	ทานอาหารไขมัน
2	SweetFood	ทานอาหารรสหวาน
4	SourFood	ทานอาหารรสเปรี้ยว
5	Bone audio	เสียงในข้อเข่า
6	Heavy lift	ยกของหนัก
7	Bend knees long time	นั่งท่าพับงอเข่านาน
8	Ground to slope	เดินขึ้นลงพื้นที่ลาดชันบ่อย
9	Exercise wrong	ออกกำลังกายผิดวิธี
10	Accidents to knees	เกิดอุบัติเหตุที่เข่า
11	Knee stiffness	ปวดหรือข้อเข่าติดขัด
12	Knee abnormal	รูปเข่าผิดปกติ
13	Knee pain	ปวดเมื่อนั่งท่าเดิมนานๆ
14	Muscle weakness	พับงอเข่านานพอียดแล้วไม่มีแรง
15	Shocks to much leg	เล่นกีฬาที่ใช้แรงกระแทกขาเยอะ
16	Pregnancy weight	รับน้ำหนักจากการตั้งครรภ์

4.3.การหาโมเดล

จากตารางตารางที่1 คือข้อมูลทั่วไป (ข้อมูลส่วนตัว)ของโมเดลจากการวิเคราะห์ผลข้อมูลผู้ที่ไม่เป็นโรคข้อเข่าเสื่อม/เป็นโรคข้อเข่าเสื่อมโดยนำโมเดลของการจำแนกข้อมูลมา 4 โมเดลเพื่อหาโมเดลที่มีประสิทธิภาพที่สุดมาเป็นโมเดลที่จะนำไปพัฒนาเป็น Web Application ซึ่งโมเดลที่เลือกมาได้แก่ Decision Tree , Naïve Bayes, Neural networks และ Sequential Minimal Optimization(SMO) จากการเปรียบเทียบผลการวิเคราะห์โมเดลของชุดข้อมูลทั้ง 4 อัลกอริทึมสรุปได้ว่า Naïve Bayes มีประสิทธิภาพดีที่สุดจึงนำโมเดลนี้มาคัดกรองผู้ที่เสี่ยงต่อการเป็นโรคข้อเข่าเสื่อมและมีระดับความเจ็บป่วยระดับใด

5.ผลการทดลอง/การทดลอง

ขั้นตอนการวิเคราะห์ความเสี่ยงของการเกิดโรคข้อเข่าเสื่อมจากข้อมูลที่เก็บมาบันทึกของกลุ่มคนที่เป็นโรคข้อเข่าเสื่อมที่ได้จากการตอบแบบสอบถามทั้งหมด 44 บั้จยจะได้อัจยที่เหมะสมเป็นบั้จยวิเคราะห์จากการทำ Attributes Selection โดยใช้เทคนิคดังนี้ อัลกอริทึม Subset selection evaluates และวิธีการค้นหาแบบ greedy hill climbing,ใน

กรณีที่ 1 คือเป็นโรคหรือไม่อยู่ 20 ปีจ้ยและ ในกรณีที่ 2 คือเป็นโรคจำแนกระดับความเจ็บปวดจำนวน 15 ปีจ้ยจากการประเมินประสิทธิภาพของโมเดลเลือกโมเดลที่เลือกมาทำการทดลองค่าพารามิเตอร์ที่ดีที่สุดสำหรับแต่ละโมเดลโดยใช้โปรแกรม WEKA แสดงผลการประเมินประสิทธิภาพของโมเดล

ผลประสิทธิภาพของแบบจำลองในแต่ละอัลกอริทึมดังตารางที่ 1 นั้นโดยการนำ Attributes จากตารางที่ 3 โดยคำนวณจากสมการการหาความน่าจะเป็นของการป่วยหรือไม่ป่วยเป็นโรคข้อเข่าเสื่อมและหาระดับความรุนแรงของโรคโดยการนำ Attributes จากตารางที่ 2 หาประสิทธิภาพของแบบจำลองในแต่ละอัลกอริทึมดังตารางที่ 3 ดังนั้นผลการวิเคราะห์ อัลกอริทึม Naïve Bays เหมาะสมที่สุดกับชุดข้อมูลทดสอบจึงใช้สมการดังต่อไปนี้นี้เป็นแบบจำลองในการวิเคราะห์เพื่อคัดกรองโรค

$$P(A|C) = P(a1|C) \times P(a2|C) \times \dots \times P(am|C) \quad (1)$$

จากสมการและข้อมูลปัจจัยได้ตัวอย่างสมการที่ใช้ในการวิเคราะห์ความเสี่ยงดังนี้

$$P(\text{Osteoarthritis}=\text{Yes}) = (\text{walking} = \text{not}|\text{Yes}) \times (\text{up stair} = \text{low}|\text{Yes}) \times (\text{sleeping} = \text{medium}|\text{Yes}) \times (\text{sit up} = \text{medium}|\text{Yes}) \times (\text{standing} = \text{medium}|\text{Yes}) \times (\text{wake up} = \text{not}|\text{Yes}) \times (\text{change stance} = \text{medium}|\text{Yes}) \times (\text{bent down} = \text{medium}|\text{Yes}) \times (\text{up down car} = \text{medium}|\text{Yes}) \times (\text{ware shoes socks} = \text{not}|\text{Yes}) \times (\text{take off shoes socks} = \text{not}|\text{Yes}) \times (\text{get out of bed} = \text{low}|\text{Yes}) \times (\text{sit toilet} = \text{low}|\text{Yes}) \times (\text{do light housework} = \text{medium}|\text{Yes}) \times (\text{Osteoarthritis} = \text{yes, medium}|\text{Yes})$$

ผลการคัดกรองจากข้อมูลของคนที่มีความเสี่ยงที่จะเป็นโรคข้อเข่าเสื่อมหรือไม่เป็นข้อเข่าเสื่อม ถ้าเป็นจะมีระดับความรุนแรงระดับใดนั้นได้ผลออกมาตามสมการที่คำนวณความน่าจะเป็นที่จะมีความเสี่ยงที่จะเป็นโรคข้อเข่าเสื่อมและมีระดับความรุนแรงตามสมการข้างต้น

5.1 ผลการทำงานของระบบตามขั้นตอน

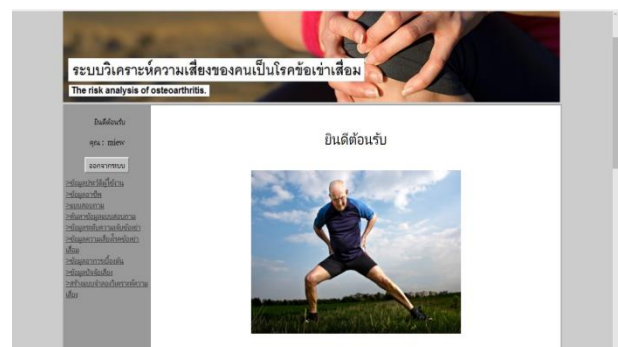
1.การเข้าสู่ระบบ

ผู้ดูแลระบบก่อนเข้าใช้งานในระบบจะต้องเข้าสู่ระบบก่อนโดยกรอกชื่อผู้ใช้งานและรหัสผ่าน จากนั้นกดที่ปุ่ม “เข้าสู่ระบบ” รูปที่ 1



รูปที่ 1 ภาพแสดงหน้าแรกและเข้าสู่ระบบ

เมื่อผู้ดูแลระบบเข้าสู่ระบบแล้ว ระบบจะแสดงหน้าแรกของผู้ดูแลระบบ ทางด้านซ้ายมือจะแสดงชื่อผู้ใช้งาน ปุ่มออกจากระบบ และแถบเมนูในส่วนของผู้ดูแลระบบ รูปที่ 2



รูปที่ 2 หน้าแรกในส่วนของผู้ดูแลระบบ

2.การใช้งานระบบส่วนของผู้ดูแลระบบ)Admin

2.1 การค้นหาข้อมูล

หน้าต่อไปนี้จะจะเป็นหน้าสำหรับการค้นหาข้อมูล สามารถค้นหาข้อมูลได้เป็นส่วนๆ ดังนี้

จากรูปที่ 2 เมื่อเลือกแถบเมนูข้อมูลประวัติผู้ใช้งานระบบจะแสดงรูปที่ 3 ซึ่งในช่องค้นหา สามารถระบุได้ว่าจะค้นหาจากคอลัมน์ใด ได้แก่ ทั้งหมด รหัสผู้ใช้งาน รหัสประวัติส่วนตัว ชื่อผู้ใช้งาน ชื่อนามสกุล ประเภทสมาชิกเบอร์โทรศัพท์ ตามข้อความที่กรอกในรหัสอีเมลตาม textbook ได้

เมื่อผู้ดูแลระบบต้องการค้นหาข้อมูลของแบบสอบถาม สามารถกดที่แถบเมนูค้นหาข้อมูลแบบสอบถาม ระบบจะแสดงรูปที่ 3

ชื่อ	รหัส	วันที่เกิด	วันที่ตรวจ	การวินิจฉัย	การให้ผล
Q010	Q010	26-Dec-16	16	ไม่	ไม่
Q012	Q012	16-Mar-17	16	ไม่	ไม่
Q018	Q018	16-Mar-17	16	ไม่	ไม่
Q019	Q019	16-Mar-17	16	ไม่	ไม่
Q017	Q017	16-Mar-17	16	ไม่	ไม่
Q017	Q017	16-Mar-17	16	ไม่	ไม่
Q017	Q017	16-Mar-17	16	ไม่	ไม่
Q017	Q017	16-Mar-17	16	ไม่	ไม่
Q017	Q017	16-Mar-17	16	ไม่	ไม่
Q017	Q017	16-Mar-17	16	ไม่	ไม่

รูปที่ 3 ภาพแสดงหน้าข้อมูลแบบสอบถาม

จากรูปที่ 6 เมื่อสมาชิกกรอกข้อมูลแบบสอบถามจนครบถ้วน ให้กดที่ปุ่ม “ไปหน้าวิเคราะห์ความเสี่ยง” ระบบจะแสดงหน้าผลการวิเคราะห์ความเสี่ยงโรคข้อเข่าเสื่อม รูปที่ 6

รูปที่ 6 ภาพแสดงหน้าผลการวิเคราะห์ความเสี่ยงโรคข้อเข่าเสื่อม

จากรูปที่ 3 ถ้าผู้ดูแลระบบต้องการดูข้อมูลประวัติส่วนตัวในส่วนหนึ่งของแบบสอบถาม

3. การใช้งานระบบส่วนของผู้ใช้ (User)

3.1 การทำแบบสอบถาม

เมื่อสมาชิกเข้าสู่ระบบแล้ว ระบบจะแสดงหน้าข้อมูลแบบสอบถาม เพื่อให้สมาชิกกรอกข้อมูลเพื่อทำการวิเคราะห์โรคข้อเข่าเสื่อม โดยจะต้องกรอกข้อมูลให้ครบทุกช่องที่กำหนด รูปที่ 4 รูปที่ 5

รูปที่ 4 ภาพแสดงหน้าแบบสอบถาม

รูปที่ 5 ภาพแสดงหน้าแบบสอบถาม

จากรูปที่ 6 เมื่อกดที่ปุ่ม “ดูผลการวิเคราะห์” ระบบจะแสดงผลการวิเคราะห์โรคข้อเข่าเสื่อมโดยระบบจะแสดงผลว่า “คุณมีความเสี่ยงโรคข้อเข่าเสื่อม” หรือ “คุณไม่มีความเสี่ยงโรคข้อเข่าเสื่อม” และระบบจะแสดงระดับการเจ็บข้อเข่าว่าสมาชิกมีความเสี่ยงโรคข้อเข่าเสื่อมอยู่ในระดับที่เท่าไร ซึ่งระดับการเจ็บข้อเข่ามีทั้งหมด 3 ระดับ คือ

ระดับที่ 1 มีการเจ็บปวดข้อเข่าน้อย ทำงานทุกอย่างได้ตามปกติ

ระดับที่ 2 มีความเจ็บปวดข้อเข่าปานกลาง ทำงานหนักไม่ได้

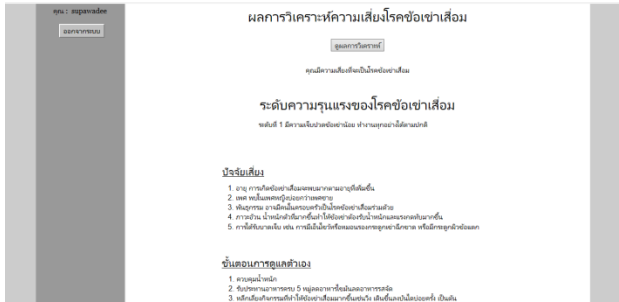
ระดับที่ 3 มีความเจ็บปวดข้อเข่ามาก เดินไม่ไหว

-ซึ่งแต่ละระดับระบบจะแสดงปัจจัยเสี่ยงและขั้นตอนการดูแลตัวเองของแต่ละระดับความเจ็บข้อเข่า รูปที่ 7 รูปที่ 8 และ รูปที่ 9 และรูปที่ 10

กรณีที่ผลวิเคราะห์เป็น “คุณไม่มีความเสี่ยงโรคข้อเข่าเสื่อม” ระบบจะไม่แสดงระดับการเจ็บข้อเข่าและจะแสดงวิธีสังเกตอาการเบื้องต้น

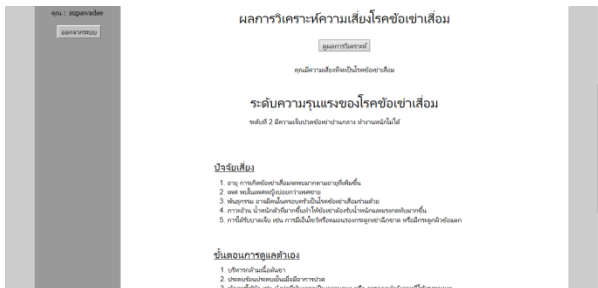
รูปที่ 7 ภาพแสดงหน้าผลการวิเคราะห์ความเสี่ยงโรคข้อเข่าเสื่อม

3.2 การดูผลการวิเคราะห์โรคข้อเข่าเสื่อม



รูปที่ 8 ภาพแสดงหน้าผลการวิเคราะห์ความเสี่ยงโรคข้อเข่าเสื่อมระดับที่

1



รูปที่ 9 ภาพแสดงหน้าผลการวิเคราะห์ความเสี่ยงโรคข้อเข่าเสื่อมระดับที่

2

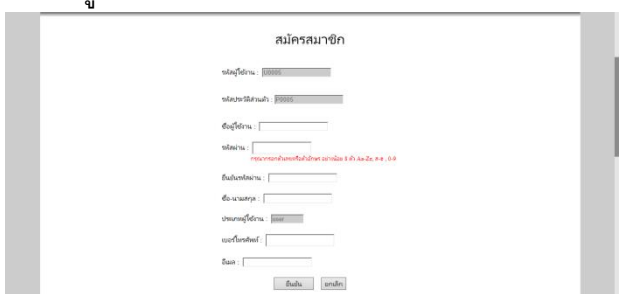


รูปที่ 10 ภาพแสดงหน้าผลการวิเคราะห์ความเสี่ยงโรคข้อเข่าเสื่อม ระดับ

ที่ 3

3.2 การสมัครสมาชิก

เมื่อผู้ใช้งานทั่วไปกดที่เมนูสมัครสมาชิกทางด้านขวามือตรงส่วนของการเข้าสู่ระบบ ระบบจะแสดงหน้าสมัครสมาชิก รูปที่ 11



รูปที่ 11 ภาพแสดงหน้าการกรอกข้อมูลการสมัครสมาชิก

6. สรุปผลการวิจัยและข้อเสนอแนะ

6.1 ผลการศึกษาพบว่า การนำเทคนิคเหมืองข้อมูลเข้ามาช่วยในการแก้ปัญหาโดยใช้ Naïve Bayes Algorithm มาใช้ในการสร้างแบบจำลองเพื่อที่จะทำการวิเคราะห์ความเสี่ยงในการเกิดโรคข้อเข่าเสื่อมเพื่อช่วยลดความเสี่ยงการเกิดโรคข้อเข่าเสื่อมในผู้สูงอายุและผู้ที่มีปัจจัยเสี่ยงให้ลดน้อยลงและเพื่อศึกษาปัจจัยเสี่ยงด้านต่างๆ พฤติกรรมที่ส่งผลต่อการเกิดโรคข้อเข่าเสื่อมโดยใช้เทคนิคการวิเคราะห์ในการจำแนกข้อมูลเข้ามาช่วยในการแก้ปัญหาซึ่งจะช่วยให้การวิเคราะห์บุคคลนั้นมีความเสี่ยงในการเกิดโรคข้อเข่าเสื่อมและยังสามารถนำแบบจำลองที่ได้ไปประยุกต์ใช้ในระบบวิเคราะห์ความเสี่ยงในการเกิดโรคข้อเข่าเสื่อมได้

6.2 ผลการวัดประสิทธิภาพของระบบจากการหาค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยสามารถสรุปได้ว่าระบบการวิเคราะห์ความเสี่ยงในการเกิดโรคข้อเข่าเสื่อมใกล้เคียงกับความจริงโดยใช้โมเดล Naïve Bayes ทั้ง 2 กลุ่มมีประสิทธิภาพค่าความถูกต้องมีค่าเท่ากับ 92.1466 % , 99.7382 % ตามลำดับ จะเห็นได้ว่าโมเดล Naïve Bayes มีประสิทธิภาพเหมาะสมกับข้อมูลที่ใช้ในการวิเคราะห์ความเสี่ยงต่อโรคข้อเข่าเสื่อม

6.3 ปัญหาการเก็บข้อมูลที่จะนำมาวิเคราะห์ความเสี่ยงในการเกิดโรคข้อเข่าเสื่อมต้องใช้เวลานานมาก ที่จะสามารถนำข้อมูลมาทำการวิเคราะห์ความเสี่ยงในการเกิดโรคข้อเข่าเสื่อมได้และยังเกิดปัญหาการเก็บข้อมูลที่เสียและไม่สามารถใช้ในการวิเคราะห์ได้จึงต้องเก็บข้อมูลเพิ่มเติมเพื่อให้ข้อมูลมีจำนวนที่สามารถใช้วิจัยทั้งสองกลุ่มตัวอย่างคือคนที่ เป็นโรคข้อเข่าเสื่อมและไม่ใช่โรคข้อเข่าเสื่อม ในขั้นตอนการเตรียมข้อมูลสำหรับการวิเคราะห์ความเสี่ยงในการเกิดโรคข้อเข่าเสื่อมตามข้อมูลส่วนตัวและพฤติกรรม รวมถึงปัจจัยเสี่ยงต่างๆ ซึ่งจะต้องเก็บข้อมูลให้ครอบคลุมชัดเจนเพื่อสามารถหาความเสี่ยงการเกิดโรคข้อเข่าเสื่อมได้

6.4 ข้อเสนอแนะสำหรับการพัฒนาในอนาคตการพิจารณาเลือกตัวแปรอาจมีการเปลี่ยนแปลงถึงผลในการเลือกตัวแปรที่เป็นปัจจัยเสี่ยงสำคัญที่ทำให้เกิดโรคข้อเข่าเสื่อมได้ การเลือกวิธีการพยากรณ์ที่มีประสิทธิภาพที่ดีกว่ามาใช้ในการสร้างแบบจำลองเพื่อนำมาวิเคราะห์ความเสี่ยงในการเกิดโรคข้อเข่าเสื่อมอาจมีการพัฒนาตามชุดข้อมูลวิจัยที่มีปัจจัยครอบคลุมรวมทั้งวิธีการที่ทันสมัยในอนาคต

7. กิตติกรรมประกาศ

ขอขอบพระคุณผู้ป่วยจากโรงพยาบาลอ่างทอง คลินิกกระดูก และข้อ ที่เสียสละเวลาช่วยทำแบบสอบถามวัดความเสี่ยงโรคข้อเข่าเสื่อมบุคคลต่างๆที่ช่วยประสานงาน จนทำให้นำข้อมูลมาช่วยในการวิเคราะห์ตัดสินใจงานวิจัยลุล่วงไปด้วยดี

8. เอกสารอ้างอิง

- [1] กิตติพล วิแสง,สิรภัทร เชี่ยวชาญวัฒนาและ คำธณ สุนันติ. "การวิเคราะห์ปัจจัยเสี่ยงของโรคเบาหวาน", The 5th National Conference on Computing and Information Technology (NCCIT2009), 2552.
- [2]รัตนาวลี ภักดีสมัย,พนิษฐา พานิชชีวะกุล. "การพัฒนาการดูแลผู้สูงอายุที่มีปัญหาปวดข้อเข่าของโรงพยาบาลส่งเสริมสุขภาพตำบลบ้านบากอำเภोजังหารจังหวัดร้อยเอ็ด",วารสารพยาบาลศาสตร์และสุขภาพ , Journal of Nursing Science & Health ปีที่ 34 ฉบับที่ 4 (ตุลาคม (ธันวาคม-Volume 34 No.4 (October-December) 2011, 2554.
- [3]นพรัตน์พจน์ จิราภรณ์.จีระศักดิ์ นำประดิษฐ์ และศักดิ์ชาย ตั้งวรรณวิทย์. "ระบบสนับสนุนทางการแพทย์เพื่อวิเคราะห์โรคหลอดเลือดแดงโป่งพองโดยใช้เทคนิคเหมืองข้อมูลกรณีศึกษาการรักษาหลอดเลือดแดงโป่งพองผ่านสายสวน", The Eleventh National Conference on Computing and Information Technology(NCCIT2015),2558.