อภินันทนาการ

สัญญาเลขที่ R2559C074

สำนักหอสมุด

รายงานวิจัยฉบับสมบูรณ์

การพัฒนาระบบจำลองพยากรณ์ปริมาณโรคอ้วนแบบอัตโนมัติผ่านอุปกรณ์พกพาเคลื่อนที่ด้วยเทคนิคฟัซซี่ดาต้าไมนิ่ง: กรณีศึกษา นักเรียนในโรงเรียนระดับชั้นประถมศึกษาในเขตจังหวัดภาคเหนือตอนล่าง (ประเทศไทย)

คณะผู้วิจัย สังกัด

ผู้ช่วยศาสตราจารย์เทวิน ธนะวงษ์ ภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ
คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร

สำนักหอสมุด มหาวิทยาลัยนเรศวร
วันลงทะเบียน - 1 มี.ค. 2562
เลขทะเบียน 10 26391
เลขเรียกหนังสือ ว QA
9
.64
ท657ร
2561

สนับสนุนโดยกองทุนวิจัยมหาวิทยาลัยนเรศวร

กิตติกรรมประกาศ

งานวิจัยฉบับนี้ได้รับเงินอุดหนุนงบประมาณรายได้ มหาวิทยาลัยนเรศวร ประจำปีงบประมาณ 2559 ได้สำเร็จลุล่วงด้วยความกรุณาอย่างยิ่งของภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร ที่เอื้อเฟื้อทรัพยากรต่างๆ ที่นำมาใช้เป็นส่วนหนึ่งของงานวิจัย จนทำให้การศึกษาวิจัยสำเร็จลุล่วงสมบูรณ์ได้ ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอขอบคุณคณะผู้บริหาร ตลอดจนเจ้าหน้าที่ฝ่ายงานวิจัย และบุคลากรของภาควิชาวิทยาการคอมพิวเตอร์ฯ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร รวมทั้งหน่วยงานสถานศึกษาของโรงเรียนระดับชั้นประถมศึกษาโดยเฉาะอย่างยิ่งของภาครัฐและองค์กรด้านสาธารณสุขต่างๆ ที่ได้ให้ความอนุเคราะห์อำนวยความสะดวกและให้ความร่วมมือ พร้อมทั้งให้คำปรึกษาในส่วนของการดำเนินการด้านเอกสารต่างๆ ให้ถูกต้องจนสำเร็จสมบูรณ์

คุณค่าและประโยชน์อันพึงมีจากการศึกษาค้นคว้าฉบับนี้ ผู้วิจัยขออุทิศแด่ผู้มีพระคุณทุกๆ ท่าน

ผู้จัดทำวิจัย

บทคัดย่อ

โรคอ้วนโดยทั่วๆไปแล้ว จะนิยมคำนวณค่าการเป็นโรคอ้วนด้วยสูตรการหาค่าดัชนีมวลกายนั้น จะพบว่ามีข้อจำกัดของการใช้งาน นั่นคือ หากเรานำเข้าข้อมูลที่มิใช่รูปแบบจำนวนเชิงปริมาณที่เป็นค่าคริสป์ (Crisp value)เพียงประเภทเดียวซึ่งเป็นที่นิยมใช้กันทั่วไป หรือแม้แต่การสร้างตัวแบบพยากรณ์โรคอ้วนที่ใช้เทคนิคเหมืองข้อมูลนั้น หากมีการนำเข้าข้อมูลของค่าน้ำหนักและค่าส่วนสูงที่เป็นปัจจัยหนึ่งสำหรับใช้เป็นค่าการพยากรณ์ให้อยู่ในลักษณะรูปแบบต่างๆ ของข้อมูล ซึ่งได้แก่ ข้อมูลในลักษณะ ค่าคริสป์ ค่าประมาณ และค่าพจน์ภาษาฟัซซี (Fuzzy Linguistic Term) เป็นต้น ก็จะไม่สามารถดำเนินการใช้อัลกอริธึมภายใต้เทคนิคเหมืองข้อมูลที่นำเข้าข้อมูลด้วยลักษณะดังกล่าวได้

งานวิจัยนี้ได้นำเสนอการแก้ไขปัญหาดังกล่าว เพื่อให้สามารถใช้งานเทคนิคเหมืองข้อมูลภายใต้ข้อจำกัดของข้อมูลนำเข้าเพื่อใช้พยากรณ์การเกิดโรคอ้วนที่มีลักษณะ ค่าคริสป์ ค่าประมาณ และค่าพจน์ภาษาฟัซซี ได้ ตามลำดับ โดยผู้วิจัยสร้างระบบฐานข้อมูลเชิงฟัซซี ในลักษณะเชิงแนวคิด (Conceptual Meta Schema) เพื่อรองรับการจัดเก็บข้อมูลที่มีลักษณะฟัซซีแอททริบิวต์ นอกจากนี้ได้ดำเนินการแมทชิ่งเชิงฟัซซีแอททริบิวต์ในลักษณะ 3 รูปแบบนั่นคือ รูปแบบที่1คือ ค่าคริสป์กับค่าพจน์ภาษาฟัซซี รูปแบบที่2 คือ ค่าประมาณกับพจน์ภาษาฟัซซี และรูปแบบสุดท้ายคือ ค่าพจน์ภาษาฟัซซีกับพจน์ภาษาฟัซซี ตามลำดับ เพื่อเลือกโหนดแอททริบิวต์จากฟัซซีแอททริบวต์เหล่านี้เพื่อนำไปใช้ในกระบวนการทำงานของเทคนิคเหมืองข้อมูลได้ ดังนั้นในงานวิจัยนี้ผู้วิจัยได้นำเสนอและเปรียบเทียบรูปแบบเชิงฟัซซีของการทำงานภายใต้อัลกอริธึมของเทคนิคเหมืองข้อมูลภายใต้รูปแบบที่เรียกว่า เทคนิคฟัซซีนิวรอลเน็ตเวิร์ค กับเทคนิคฟัซซีต้นไม้ช่วยตัดสินใจเพื่อคัดเลือกการทำงานของเทคนิคที่เหมาะสมสำหรับใช้สร้างตัวแบบจำลองพยากรณ์โรคอ้วนให้มีความหลากหลายจากรูปแบบของข้อมูลนำเข้าเพื่อใช้สำหรับการพยากรณ์

คำสำคัญ: โรคอ้วน, ฟัซซีนิวรอลเน็ต, ฟัซซีต้นไม้ช่วยตัดสินใจ

Abstract

In general, obesity is calculated using the BMI, which is found to have limitations of use when we insert only one type of non-quantitative data as a crisp value, which is commonly used, or even create an obesity forecasting model using data mining techniques. If there is an input of weight and height data as a forecasting factor in various types of data, such as crisp value, estimated value and Fuzzy Linguistic Term value, etc. The algorithm will not possibly work under the data mining technique with such input data.

This research presented the solution for this problem in order to enable data mining techniques under limitations of inserted data, including crisp value, estimated value and fuzzy linguistic term value, for forecasting obesity. The researchers created a fuzzy database system in the form of Conceptual Meta Schema to support fuzzy attributed data storage. In addition, there were 3 types of fuzzy attribute matching as follows; Type 1 was matching crisp value with fuzzy linguistic term value, Type 2 was matching estimated value with fuzzy linguistic term value, and Type 3 was matching fuzzy linguistic term value with fuzzy

linguistic term value, respectively. The attribute node from these fuzzy attributes would be selected to be used in the work of data mining techniques. Therefore, in this research, the researchers presented and compared the fuzzy model of working under the algorithm of data mining techniques in the form called Fuzzy Neural Network Techniques and Fuzzy Decision Tree Techniques to help deciding and selecting the appropriate techniques for creating obesity forecasting model with diversity of input data formats used for forecasting.

Key words: Obesity, Fuzzy Neural Net, Fuzzy Decision Tree

สารบัญเรื่อง

หน้า

บทที่1 บทนำ (Introduction)........8
1.1 ที่มาของปัญหา........8
1.2 วัตถุประสงค์........9
1.3 ขอบเขตการวิจัย........9
1.4 วิธีการดำเนินการวิจัยและระยะเวลาทำวิจัย........9
1.5 ประโยชน์ที่คาดว่าจะได้รับ........10

บทที่2 งานวิจัยที่เกี่ยวข้องและเอกสารทางทฤษฎี........11
2.1 โรคอ้วน........11
2.2 การทำเหมืองข้อมูล........11
2.2.1 ต้นไม้ตัดสินใจ (Decision Tree)........12
2.2.2 โครงข่ายประสาทเทียม (Neural Network)........12
2.3 โปรแกรมWEKA........14
2.4 ตรรกศาสตร์ความคลุมเครือ (Fuzzy logic)........15
2.5 การทบทวนวรรณกรรม/สารสนเทศ (Information) ที่เกี่ยวข้อง........16

บทที่3 วิธีดำเนินการวิจัย........18
3.1สถาปัตยกรรมของระบบการทำงานในงานวิจัย........18
3.2 การรวบรวมปัจจัยที่เกี่ยวข้องของการสร้างแบบจำลองพยากรณ์โรคอ้วน........18
3.3 การคัดเลือกปัจจัยนำเข้าด้วยการวิเคราะห์
สมการสหสัมพันธ์เพียร์สัน (Pearson's correlation analysis)........20
3.4 การออกแบบระบบจัดการฐานข้อมูล........22
3.4.1 การออกแบบความสัมพันธ์ของข้อมูลในระดับเชิงแนวคิด........22
3.5 การสร้างแบบจำลองพยากรณ์โรคอ้วนด้วยเทคนิคฟัซซีนิวรอลเน็ตเวิร์ค
กับฟัซซีต้นไม้ช่วยตัดสินใจ........24
3.5.1 การสร้างแบบจำลองด้วยเทคนิคฟัซซีนิวรอลเน็ตเวิร์ค........25
3.5.2 การสร้างแบบจำลองด้วยเทคนิคฟัซซีต้นไม้ช่วยตัดสินใจ........26
3.6 การนำเสนอ Fuzzy data mining technique........27
3.7 หลักการทำงานของ Fuzzy neural network และ Fuzzy decision trees........28
3.8 การนำเสนอค่าฟัซซีแอททริบิวต์ (Fuzzy attribute value representation)........29
3.9 การทำงานของ Fuzzy Attribute Matching Techniques........30

บทที่4 ผลการทดลอง..........32
4.1 ผลลัพธ์การพยากรณ์โรคอ้วนจากข้อมูลนำเข้าภายใต้การใช้เทคนิคฟัซซีดาต้าไมนิง..........33
4.2 ตัวอย่างการทำงานของแอพพลิเคชันที่ได้พัฒนา..........36

สรุปผลการวิเคราะห์และวิจารณ์..........39
บรรณานุกรม..........40
ภาคผนวก ก..........41
ภาคผนวก ข..........49
ภาคผนวก ค..........50
ภาคผนวก ง..........51
ภาคผนวก จ..........55

มหาวิทยาลัยนเรศวร

บทที่1
บทนำ (Introduction)

1.1 ที่มาของปัญหา

ในปัจจุบันการบริโภคอาหารของบุคคลทั่วๆ ไปในสังคมปัจจุบันนั้น มีความแตกต่างจากในอดีตมาก ทั้งนี้เนื่องมาจากการแข่งขันในการประกอบอาชีพและการดำรงชีวิตในสังคมที่มีความเจริญทั้งทางด้านเศรษฐกิจและด้านเทคโนโลยี จึงต้องการความรวดเร็วในการบริโภคอาหาร ทำให้เกิดปัญหาส่งผลกระทบต่อสุขภาพของบุคลในแต่ละครอบครัวมากยิ่งขึ้น รวมทั้งในปัจจุบันนั้น สื่อเทคโนโลยี หรือสิ่งพิมพ์ต่างๆ ได้เข้ามาเกี่ยวข้องเกี่ยวและยั่วยุทั้งด้านการลด แลก แจก แถมสินค้าต่างๆ เกินความจำเป็นในการบริโภคของบุคคลทั่วๆ ไป ดังนั้น คนในสังคมโดยเฉพาะเด็กที่ไม่สามารถแยกแยะหรือตัดสินใจในเลือกบริโภคอาหารต่างๆ ที่เป็นประโยชน์ต่อสุขภาพได้นั้น ยิ่งมีโอกาสเสี่ยงที่จะมีปัญหาทางสุขภาพที่สูงมากขึ้น จากปัญหาที่ได้กล่าวไว้ข้างต้นจากการบริโภคอาหารนั้น ทำให้เกิดโรคอ้วน ซึ่งปัญหาโรคอ้วนที่เกิดขึ้นในเด็กนั้น ถือได้ว่าเป็นปัญหาใหญ่ที่หลายองค์กรให้ความสำคัญ เช่น สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) กระทรวงศึกษาธิการ ได้จัดทำโครงการเด็กไทยสุขภาพดี เน้นชูนวัตกรรมสื่อการเรียนการสอนยุคใหม่ ที่สอดแทรกสาระความรู้ความบันเทิงเรื่องสุขภาพ และโภชนาการ โดยพัฒนาวิธีการนำเสนอสื่อการเรียนการสอนสมัยใหม่ เผยแพร่ไปยังโรงเรียนระดับประถมศึกษา มากกว่า 2,000 แห่งทั่วประเทศ รวมทั้งการสร้างจิตสำนึกให้กับครู และผู้ปกครองได้ตระหนักถึงความสำคัญในการดูแลสุขภาพเด็กตั้งแต่เยาว์วัย อย่างไรก็ตามปัญหาของโรคอ้วนของเด็กนั้นยังเป็นปัญหารุนแรงอยู่ในทุกประเทศ ซึ่งแม้แต่ในประเทศไทยนั้น ปัญหาที่เกิดจากโรคอ้วนของเด็กส่งผลกระทบทำให้หน่วยงานของรัฐสูญเสียงบประมาณจำนวนมากเพื่อนำไปใช้ในการวางมาตรการและรณรงค์ รวมทั้งการส่งเสริมสุขภาพอนามัยที่ถูกต้องเพื่อป้องกันการเกิดโรคอ้วนในเด็ก ที่จะก่อให้เกิดปัญหาทรัพยากรบุคคลของประเทศขาดคุณภาพได้ในอนาคต รวมทั้งมีโอกาสการเกิดโรคแทรกซ้อนต่างๆ อันผลเนื่องมาจากการเป็นโรคอ้วน สำหรับปัญหาของจำนวนการเกิดโรคอ้วนในเด็กนั้น ในปัจจุบันหน่วยงานของภาครัฐต่างๆ ได้ดำเนินการภายใต้การจัดเก็บข้อมูลที่ไม่ชัดเจนโดยอาศัยเพียงการจัดเก็บข้อมูลภายใต้ในการตรวจทางการแพทย์ที่จำเป็นต้องได้รับการตรวจโดยตรง จึงทำให้เกิดความยุ่งยาก และเสียเวลาในการปฏิบัติงานมาก อย่างไรก็ตาม ข้อมูลจำนวนการเกิดโรคอ้วนที่มีการจัดเก็บไว้จำนวนมากในแต่ละปีเหล่านั้น มีประโยชน์อย่างมากหากนำไปประระยุกต์ใช้ร่วมกับเทคโนโลยีและหลักทางวิชาการเพื่อสร้างแบบระบบจำลองสำหรับใช้พยากรณ์การเกิดโรคอ้วนของนักเรียนโดยเฉพาะอย่างยิ่งนักเรียนในระดับชั้นประถมศึกษาในเขตจังหวัดภาคเหนือตอนล่างที่เป็นพื้นที่เบื้องต้นของการทำวิจัยในครั้งนี้ เพื่อทำการคาดการณ์จำนวนการเกิดโรคอ้วนของเด็กในโรงเรียนต่างๆ ได้อย่างแม่นยำมากขึ้น รวมทั้งสามารถใช้เป็นองค์ความรู้เพื่อป้องกันการเกิดโรคอ้วนของเด็กได้ภายใต้ข้อแนะนำจากปัจจัยที่ควรหลีกเลี่ยงจากสภาพแวดล้อมรอบตัวของเด็ก ทั้งนี้จำเป็นที่จะต้องอาศัยผู้ปกครอง และครูคอยชี้แนะ และเป็นผู้ที่จะนำแอพพลิเคชั่นที่ได้จากงานวิจัยนี้ไปใช้ เพื่อใช้เป็นเครื่องมือหนึ่งสำหรับสร้างแนวคิดที่จะดำเนินการปรับเปลี่ยนสภาพแวดล้อมให้เหมาะสมกับเด็กโดยเฉพาะอย่างยิ่งในเขตโรงเรียน

ดังนั้น ผู้วิจัยจึงเล็งเห็นความสำคัญที่จะประยุกต์ใช้เทคโนโลยีสารสนเทศเพื่อสร้างโมเดลในการพยากรณ์และวิเคราะห์ปริมาณของเด็กนักเรียนชั้นประถมศึกษาที่มีโอกาสเป็นโรคอ้วน โดยผลงานวิจัยครั้งนี้สามารถนำไปใช้ให้เกิดประโยชน์ต่อการคาดการณ์ของปริมาณการเกิดโรคอ้วนของนักเรียนระดับชั้นประถมศึกษาในเขตภาคเหนือตอนล่าง ซึ่งเป็นพื้นที่เบื้องต้นของการทำวิจัย ทั้งนี้ผลที่ได้จากแบบจำลองที่สามารถจะทำการคำนวณผลได้แม่นยำและมีระดับความน่าเชื่อถือได้นั้นช่วยในการกำหนดแนวทางป้องกัน

และวางแผนควบคุมปริมาณการเกิดของโรคอ้วนที่เกิดและเป็นปัญหาที่ควรเร่งแก้ไขของภาครัฐภายใต้หน่วยงานที่เกี่ยวข้องอย่างจริงจัง เพราะฉะนั้น ผู้วิจัยจึงสนใจที่จะนำเสนองานวิจัย ภายใต้หัวข้อเรื่องการพัฒนาระบบจำลองพยากรณ์ปริมาณโรคอ้วนแบบอัตโนมัติผ่านอุปกรณ์พกพาเคลื่อนที่ด้วยเทคนิคฟัซซี่ดาต้าไมนิ่ง: กรณีศึกษา นักเรียนในโรงเรียนระดับชั้นประถมศึกษาในเขตจังหวัดภาคเหนือตอนล่าง ทั้งนี้ในการดำเนินการวิจัยจะประกอบไปด้วย ส่วนต่างๆ นั่นคือ ส่วนของการจัดเตรียมข้อมูลองค์ความรู้และปัจจัยพื้นฐานที่ส่งผลต่อการเกิดโรคอ้วน การจัดเตรียมข้อมูลเชิงฟัซซีของการเกิดโรคอ้วนในรูปแบบไฟล์ .ARFF ภายใต้ลักษณะของ Training data sets ส่วนของการจัดจำแนกข้อมูลภายใต้พื้นฐานเทคนิคการจำแนกข้อมูล ด้วยอัลกอริธึมที่ได้จากโมเดลของโครงข่ายประสาทเทียม และอัลกอริธึมจากการใช้โมเดลของต้นไม้ตัดสินใจ พร้อมทั้งทำการเปรียบเทียบและเลือกโมเดลที่เหมาะสมในการทำงานของระบบที่สอดคล้องกับข้อมูลนำเข้าและพัฒนาแอพพลิเคชั่นบนอุปกรณ์มือถือที่สามารถทำงานภายใต้อัลกอริธึมที่เหมาะสมสำหรับพยากรณ์ปริมาณการเกิดโรคอ้วน

1.2 วัตถุประสงค์

1) เพื่อใช้เทคนิคเหมืองข้อมูลวิเคราะห์และพยากรณ์ปริมาณการเกิดโรคอ้วนของนักเรียนระดับชั้นประถมศึกษาในเขต 8 จังหวัดภาคเหนือตอนล่าง
2) เพื่อศึกษาและเปรียบเทียบการใช้อัลกอริธึมโครงข่ายประสาทเทียมกับต้นไม้ตัดสินใจที่เหมาะสมสำหรับวิเคราะห์และพยากรณ์ปริมาณการเกิดโรคอ้วนผ่านโปรแกรมWEKA
3) เพื่อพัฒนาแอพพลิเคชั่นสำหรับใช้พยากรณ์การเกิดโรคอ้วนแบบอัตโนมัติบนอุปกรณ์มือถือลักษณะออนไลน์ ที่หน่วยงานทางการศึกษาและสาธารณสุขสามารถนำไปใช้พยากรณ์ปริมาณการเกิดโรคอ้วนของนักเรียนระดับชั้นประถมศึกษาได้

1.3 ขอบเขตการวิจัย

1) รวบรวมข้อมูลเบื้องต้นจากนักเรีนระดับชั้นประถมศึกษาในเขตภาคเหนือตอนล่างจำนวน 965 ตัวอย่าง
2) ออกแบบระบบฐานข้อมูลในระดับเชิงแนวคิดด้วยเทคนิคไนแอม (NIAM) และจัดเก็บข้อมูลในลักษณะระบบจัดการฐานข้อมูลเชิงฟัซซี
3) คัดเลือกแอททริวิต์ที่เหมาะสมด้วยเทคนิคการหาความสัมพันธ์เพียร์สัน ระหว่างตัวแปรตามกับตัวแปรพยากรณ์
4) สร้างแบบจำลองภายใต้การกำหนดแอททริบิวต์เชิงฟัซซี 3 แอททริบิวต์ได้แก่ น้ำหนัก ส่วนสูงและ รายได้ของครอบครัว พร้อมเปรียบเทียบการทำงานของเทคนิคฟัวซีนิวรอลเน็ต กับฟัซซีต้นไม่ช่วยตัดสินใจ
5) พัฒนาแอปพลิเคชันระบบตรวจพยากรณ์โรคอ้วนจากแบบจำลองที่เหมาะสม

1.4 วิธีการดำเนินการวิจัย และ ระยะเวลาทำการวิจัย

1) ศึกษาค้นคว้าทางทฤษฎีและรวบรวมปัจจัยที่เกี่ยวข้องของการเกิดโรคอ้วน
2) รวบรวมข้อมูลและองค์ความรู้โรคอ้วนจากหน่วยงานสาธารณสุข และหน่วยงานอื่นๆ ที่เกี่ยวข้องกับการดำเนินการวิจัย
3) ดำเนินการจำแนกและวิเคราะห์ความสัมพันธ์ของข้อมูลเพื่อใช้ในการจัดเตรียมข้อมูลนำเข้าสู่

ระบบ และการจัดเก็บข้อมูลเพิ่มเติมเพื่อความสมบูรณ์ของการดำเนินการวิจัย

4) การสร้างโมเดลระบบและการสอนข้อมูล โดยการสร้างโมเดลต้นไม้ตัดสินใจ และโมเดลโครงข่ายประสาทเทียม

5) พัฒนาแอพพลิเคชั่นสำหรับใช้พยากรณ์ปริมาณการเกิดโรคอ้วนในนักเรียนระดับชั้นประถมศึกษาเขตจังหวัดภาคเหนือตอนล่าง

6) เขียนรายงาน นำเสนอผลงานวิจัย จัดพิมพ์รูปเล่มพร้อมส่งผลงานที่เสร็จสมบูรณ์

ตารางแสดงแผนงานและผลงาน (ในแต่ละช่วงเวลา) นับตั้งแต่เริ่มโครงการจนถึงสิ้นสุดโครงการ

กิจกรรม	เดือนที่											
	1	2	3	4	5	6	7	8	9	10	11	12
1.ค้นคว้าบทความ และทฤษฎีที่เกี่ยวข้อง	√	√										
2.เก็บรวบรวมข้อมูลภาค สนาม และในหน่วย งานต่างๆ ที่เกี่ยวข้อง		√	√	√	√	√	√					
3. จำแนกและวิเคราะห์ความสัมพันธ์ของข้อมูลเพื่อใช้ในการจัดเตรียมข้อมูลนำเข้าสู่ระบบ			√	√	√	√	√	√				
4. สร้างโมเดลระบบและการสอนข้อมูล โมเดลต้นไม้ตัดสินใจ โมเดลโครงข่ายประสาทเทียม และเปรียบเทียบประสิทธิภาพการทำงาน					√	√	√	√	√	√		
5.พัฒนาแอพพลิเคชั่นสำหรับใช้พยากรณ์ปริมาณการเกิดโรคอ้วน และทดสอบการทำงานพร้อมทั้งการปรับปรุงแก้ไขผลงาน					√	√	√	√	√	√	√	
6. เขียนรายงาน นำเสนอผลงานวิจัย จัดพิมพ์รูปเล่มพร้อมส่งผลงานที่เสร็จสมบูรณ์										√	√	√

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1) ได้ประยุกต์ใช้เทคนิคเหมืองข้อมูลสร้างแบบจำลองระบบพยากรณ์ปริมาณการเกิดโรคอ้วนในนักเรียนระดับชั้นประถมศึกษาเขตจังหวัดภาคเหนือตอนล่าง

2) ได้แอพพลิเคชั่นบนอุปกรณ์มือถือที่สามารถทำงานภายใต้อัลกอริธึมที่เหมาะสมสำหรับพยากรณ์ปริมาณการเกิดโรคอ้วน

3) ได้ประยุกต์ใช้เทคโนโลยีสำหรับเป็นเครื่องมือต่อภาครัฐทั้งทางด้านการศึกษาและสาธารณสุขได้ตระหนักถึงสุขภาพของเด็กนักเรียนระดับชั้นประถมศึกษาให้มากยิ่งขึ้น

บทที่2
งานวิจัยที่เกี่ยวข้องและเอกสารทางทฤษฎี

2.1 โรคอ้วน

ความเข้าใจเกี่ยวกับโรคอ้วน [12] ในทางวิชาการกับความเข้าใจของบุคคลทั่วๆ ไปนั้น มีความแตกต่างกัน นั่นคือความวิตกกังวลของคนทั่วๆ ไปว่าตนเอง อ้วน ซึ่งในความเป็นจริงแล้วน้ำหนักของบุคคลเหล่านั้นยังอยู่ในเกณฑ์ปกติ ซึ่งในทางวิชาการนั้นจะมีเกณฑ์ที่ใช้ในการวินิจฉัยว่าเป็นโรคอ้วนหรือไม่ ตัวบ่งชี้ที่ใช้ ในการประเมินโรคอ้วน ได้แก่ ดัชนีมวลกาย รอบเอว สัดส่วนรอบเอวต่อรอบสะโพก สัดส่วนรอบเอวต่อส่วนสูง และปริมาณร้อยละของไขมันในร่างกาย บุคลากรหรือนักวิจัยมักจะไม่แน่ใจว่าควรเลือกใช้ตัวบ่งชี้ใดและเลือกช่วงค่าของตัวบ่งชี้ในช่วงค่าใดจึงจะคัดกรองโรคอ้วนได้ดี โดยนิยมใช้ดัชนีมวลกาย และรอบเอว มีการใช้สัดส่วนรอบเอวต่อรอบสะโพก และสัดส่วนรอบเอวต่อส่วนสูงอยู่บ้าง แต่มีการใช้ร้อยละของไขมันในร่างกายค่อนข้างน้อย ซึ่งหากคำนวณโดยใช้สูตรหาค่าดัชนีมวลกายดังข้างล่างนี้ (ดังสมการ 1) ควรมีค่าต่ำกว่า 23 กก/ม^2 หรืออาจดูความสัมพันธ์ระหว่างน้ำหนักและส่วนสูงโดยใช้กราฟอ้างอิงการเจริญเติบโต ไม่ควรเกิน 90 นอกจากนี้รายละเอียดส่วนสูงและน้ำน้ำหนักของระดับนักเรียนชั้นประถมศึกษาปีที่ 1-6 จะแสดงถึงสัดส่วนของค่ามาตรฐานที่ถูกใช้เป็นเกรณฑ์ในการวัดความอ้วน (ดังตารางที่ 2.1) มีรายละเอียดดังต่อไปนี้

$$\text{ดัชนีมวลกาย (BMI)} = \frac{\text{น้ำหนักตัว (หน่วยกิโลกรัม)}}{\text{ความสูง}^2\ \text{(หน่วยเมตร}^2\text{)}} \qquad (1)$$

ตารางที่ 2.1 แสดงน้ำหนักและส่วนสูงที่ควรจะเป็นในช่วงอายุตามเกณฑ์มาตรฐานการเจริญเติบโตของเพศชายและเพศหญิง (แหล่งข้อมูล: กรมอนามัย กระทรวงสาธารณสุข 2543)

อายุ (ปี)	เพศชาย		เพศหญิง	
	น้ำหนัก (กก.)	ส่วนสูง (ซม.)	น้ำหนัก (กก.)	ส่วนสูง (ซม.)
6	16.6-25.4	107.7-121.3	16.1-24.7	107.4-120.8
7	18.3-28.8	112.8-127.4	17.7-28.7	112.4-126.8
8	20.0-32.2	117.4-133.2	19.3-32.5	117.0-132.4
9	21.5-36.6	121.8-138.3	21.2-37.4	121.9-139.1
10	23.6-40.8	126.2-143.4	23.4-42.1	127.1-146.1
11	25.6-45.2	130.5-149.4	26.1-46.5	132.9-152.6
12	28.1-50.0	135.1-156.9	29.4-50.2	138.8-156.9
13	31.6-51.6	140.9-164.4	33.0-53.1	143.5-160.2

2.2 การทำเหมืองข้อมูล

การทำเหมืองข้อมูล [1][10] สามารถวิเคราะห์ข้อมูลได้จริงตามแนวคิด (Concept) ที่วางไว้นั้นมีอยู่ด้วยกันหลายอย่าง เทคนิคแต่ละอย่างก็เหมาะสมกับการวิเคราะห์ข้อมูลทางด้านธุรกิจแตกต่างกันออกไปอย่างไรก็ตาม โดยทั่วไปแล้ว เทคนิคที่นำมาใช้ส่วนใหญ่ มี 4 ประเภท การ Classification การ Clustering

การ Association และการทำVisualization ตามลำดับ สำหรับตัวอย่างของการใช้เทคนิคด้าน classification นั้นสามารถอธิบายความสำคัญได้นั่นคือ

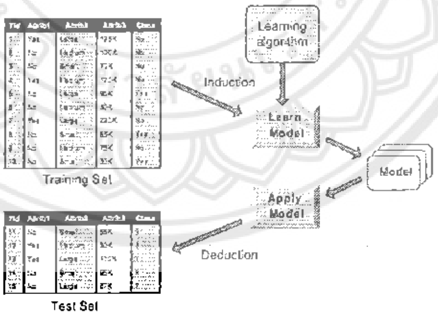
- เป็นเทคนิคในการจำแนกกลุ่มข้อมูลด้วยคุณลักษณะต่างๆที่ได้มีการกำหนดไว้แล้ว เทคนิคนี้เหมาะกับการสร้างแบบจำลองเพื่อการพยากรณ์ค่าข้อมูล (Predictive Modeling) ในอนาคตจากการที่ได้จำแนกกลุ่มข้อมูลตัวอย่างได้แล้ว ซึ่งลักษณะนี้เรียกว่า “Supervised Learning”
- เทคนิคการClassification แบ่งออกเป็น 2 รูปแบบคือ Tree Induction และ Neural Induction

2.2.1 ต้นไม้ตัดสินใจ (Decision Tree)

เป็นการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบโครงสร้างต้นไม้(Decision Trees) [9] ซึ่ง Decision Trees จะมีการทำงานแบบ Supervised Learning (คือการเรียนรู้ แบบมีครูสอน) สามารถสร้างแบบจำลองการจัดหมวดหมู่ได้จากกลุ่มตัวอย่างข้อมูลที่กำหนดไว้ก่อนล่วงหน้า เรียกว่า Training set ได้อัตโนมัติ และพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้ด้วยรูปแบบของ Tree โครงสร้างประกอบด้วย Root Node, Child และ Leaf Node กระบวนการทำ tree decision มีขั้นตอนการทำงานดังต่อไปนี้ คือ (ดังภาพที่2.1)

1). แบ่งข้อมูลตัวอย่าง (Samples Data) ออกเป็น 3 ส่วนได้แก่
- Training Datasets
- Validation Datasets
- Test Datasets

2). นำ Training Datasets มาสร้าง Decision Tree
3). ใช้ Validation Datasets วัดความถูกต้องในการจำแนกของ Tree ที่สร้าง
4). ทำซ้ำข้อ 2,3 เพื่อให้ได้ความถูกต้องสูงสุด
5). ใช้ Testing Datasets มาสอบทดกับ Tree ที่ได้เพื่อวัดความถูกต้อง

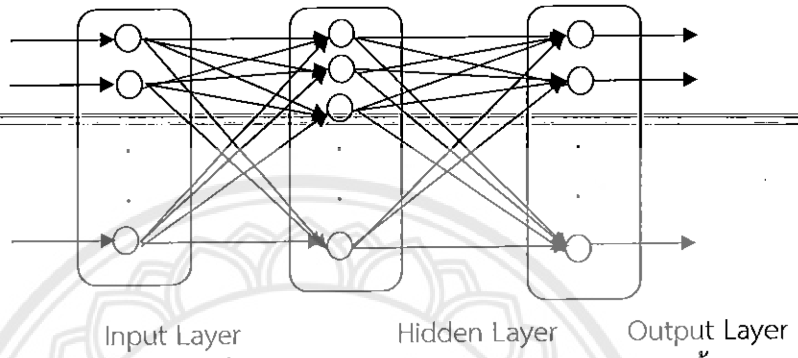


ภาพที่ 2.1 แสดงการกำหนด training data sets และการ classification ด้วย tree decision (แหล่งที่มา: Introduction to Data Mining, by Tan, Steinbach, Kumar)

2.2.2 โครงข่ายประสาทเทียม (Neural Network)

แบบจำลองระบบโครงข่ายประสาทเทียม[3][8] เป็นแบบจำลองทางคณิตศาสตร์ที่ประกอบไปด้วยโครงสร้างการทำงาน และการเชื่อมต่อเลียนแบบระบบการ

ทำงานของเซลล์สมองของมนุษย์ งานวิจัยด้านโครงข่ายประสาทเทียมมีความแพร่หลายอย่างมากที่ใช้กระบวนการแพร่กลับ (Back Propagation Algorithm) โดยทั่วไปแล้วโครงข่ายของแบบจำลองประกอบด้วยชั้นของ node 3 ชั้น ได้แก่ ชั้นข้อมูลด้านเข้า (Input layer) ชั้นแฝง (Hidden layer) และชั้นแสดงผล (Output layer) ข้อมูลจากชั้นข้อมูลด้านเข้าจะถูกคำนวณและส่งผ่านฟังก์ชั่นแปลงรูปจากชั้นแฝงไปยังชั้นแสดงผล โดยหลักการของกระบวนการเรียนรู้ของแบบจำลองคือ การเปลี่ยนแปลงค่าถ่วงน้ำหนักของแต่ละการเชื่อมต่อ เพื่อปรับผลลัพธ์ของแบบจำลองให้ใกล้เคียงกับค่าความจริงมากที่สุดโดยอาศัยการแพร่กลับ (ดังภาพที่2.2)



ภาพที่ 2.2 แสดงโครงข่ายประสาทเทียมแบบหลายชั้น

(แหล่งที่มา: Holland. 1975. Adaptation in Natural and Artificial System)

- ชั้นอินพุต (Input Layer)
  - เป็นชั้นที่รอรับข้อมูลเข้าสู่โครงข่ายประสาทเทียม
  - ข้อมูลที่นำเข้าสู่โครงข่าย ก็คือข้อมูลที่ผ่านการเตรียมข้อมูลดังที่ได้แสดงไว้ในขั้นตอนการเตรียมข้อมูลของโครงข่ายประสาทเทียม
- ชั้นซ่อน (Hidden Layer)
  - เป็นชั้นที่เพิ่มประสิทธิภาพในการจัดกลุ่มข้อมูล เพื่อที่จะส่งต่อไปยังชั้นเอาต์พุต
- ชั้นเอาต์พุต (Output Layer)
  - เป็นชั้นที่คำนวณหาผลลัพธ์ของโครงข่ายประสาทเทียม
  - จำนวนโหนดในชั้นเอาต์พุตจะขึ้นอยู่กับจำนวนกลุ่มข้อมูลที่ต้องการจัด โดยมีสมการในการคำนวณดังสมการ $f(\sum_{j=1}^{m} y_j w_{jk} + \theta_k)$

โดยที่ Z_k คือ ผลลัพธ์ในชั้นเอาต์พุตโหนดที่ *k*

y_j คือ ผลลัพธ์ในชั้นซ่อน หรือข้อมูลส่งออกในชั้นซ่อนโหนดที่ *j*

w_{jk} คือ น้ำหนักบนเส้นเชื่อมระหว่างโหนดที่*j*ในชั้นซ่อนและโหนดที่ *k* ในชั้นเอาท์พุต

θ_k คือ ค่าไบแอสของโหนดที่ k ในชั้นเอาท์พุต

m คือ จำนวนโหนดทั้งหมดของชั้นซ่อน

$f(n)$ คือ Sigmoid Function

ซึ่ง Sigmoid Function หาได้จากสมการที่นำไปคิดดังนี้

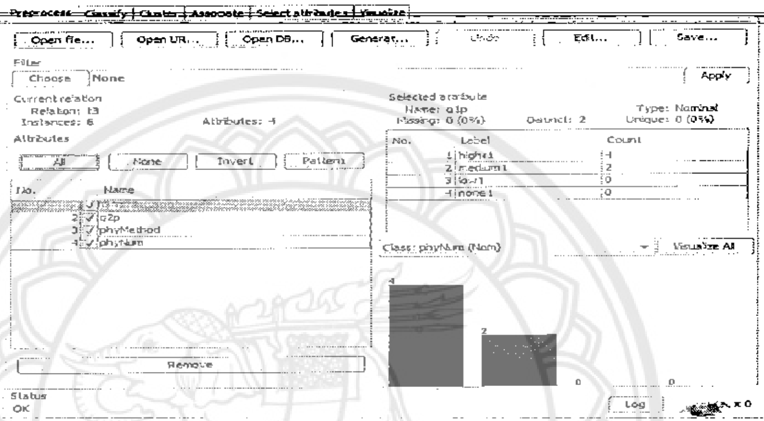
$$f(u) = \frac{1}{1+e^{-u}}$$

โดยที่ u คือ ผลลัพธ์ที่ได้จาก $\sum_{i=1}^{n} x_i w_{ij} + \theta_j$

e มีค่าประมาณ 2.71828

2.3 โปรแกรมWEKA

เวก้า [5] เป็นโปรแกรมคอมพิวเตอร์แบบโอเพ่นซอร์ส ที่สามารถนำมาใช้ได้โดยไม่เสียค่าลิขสิทธิ์ของการใช้โปรแกรม ซึ่งเป็นโปรแกรมที่สำคัญตัวหนึ่งในการนิยมนำมาใช้ทำเหมืองข้อมูล เพื่อการพยากรณ์หรือวิเคราะห์ข้อมูล เป็นต้น ซึ่งโปรแกรมนี้ถูกพัฒนาขึ้นโดยมหาวิทยาลัย WAIATO ประเทศนิวซีแลนด์ โดยสามารถดาวน์โหลดได้จากเว็บไซต์ www.cs.waikato.ac.nz/ml/weka (ดังภาพที่2.3)



ภาพที่ 2.3 แสดงการใช้งานโปรแกรมเวก้า

- Explorer เป็นโปรแกรมที่ออกแบบในลักษณะ GUI
- Experimenter เป็นโปรแกรมที่ออกแบบการทดลองและการทดสอบผล
- KnowledgeFlow เป็นโปรแกรมออกแบบผังการไหลของความรู้
- Simple CLI (Command Line Interface) เป็นโปรแกรมรับคำสั่งการทำงานผ่านการพิมพ์

รายละเอียดภายใต้เมนูหลักของ Explorer

- Preprocess การเตรียมข้อมูล
- Classify เป็นโมดูลการทำเหมืองข้อมูลแบบการจำแนกประเภท (Classification)
- Cluster เป็นโมดูลการทำเหมืองข้อมูลแบบการแบ่งกลุ่ม (Clustering)
- Associate เป็นโมดูลการทำเหมืองข้อมูลแบบกฎความสัมพันธ์ (Association rule)
- Select attribute เป็นโมดูลการเลือกลักษณะเฉพาะ
- Visualize นำเสนอข้อมูลด้วยภาพนามธรรมสองมิติ

การเตรียมข้อมูลเพื่อทำการวิเคราะห์ด้วยโปรแกรม WEKA

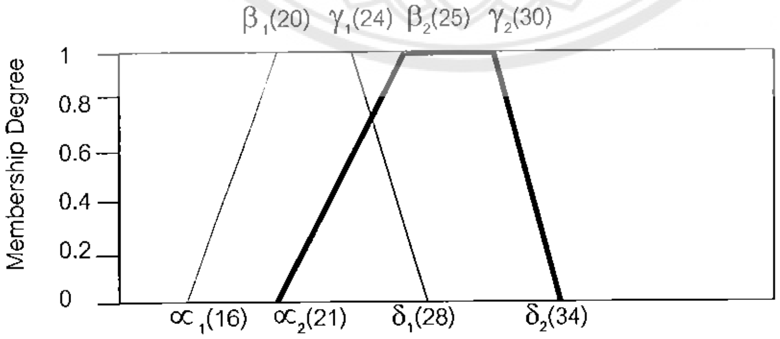
สำหรับการทำงานของโปรแกรม WEKA นั้น จะสนับสนุนการทำงานของแฟ้มข้อมูลในรูปแบบของ ASCII ได้แก่ ไฟล์นามสกุล.csv ไฟล์นามสกุล.arff ไฟล์นามสกุล .c45 หรือการอ่านจากแฟ้มข้อมูลจากเครือข่ายผ่าน URL หรือการใช้ฐานข้อมูลที่เชื่อมโยงผ่านไดร์เวอร์ JDBC เป็นต้น

2.4 ตรรกศาสตร์ความคลุมเครือ (Fuzzy logic)

จากการอธิบายปัญหา (Problem description) โดยทั่วๆ ไปนั้น เราจะพบว่าในบางครั้งปัญหาอุปสรรคที่เกิดขึ้นของการอธิบายปัญหา ก็คือ ความคลุมเครือ (Fuzzy)ที่เกิดขึ้น ตัวอย่างเช่น "ระดับอาการอ้วนมาก" หรือ "รับประทานอาหารปริมาณปานกลาง" เป็นต้น จากสองประโยคข้างต้นภายใต้คีย์เวิร์ดของ *"อ้วนมาก"* และ *"ปริมาณปานกลาง"* เราสามารถที่จะนำหลักการทำงานของ ฟัซซีลอจิก (fuzzy logic) [7][11] ซึ่งเป็นเครื่องมือที่ช่วยในการตัดสินใจภายใต้ความไม่แน่นอนหรือคลุมเครือของข้อมูลมาประยุกต์ใช้งานร่วมกันได้ ซึ่งโดยทั่วไปแล้วได้มีการนำหลักการทำงานของฟัซซีเซตใช้ในการแทนแนวคิดของภาษา ดังเช่น 'Very small', 'Small', 'Medium' และอื่นๆ เราเรียกคำเหล่านี้ว่าค่าของตัวแปรภาษา (linguistic value) หรือพจน์ภาษา (linguistic term) ลักษณะการเปรียบเทียบเท่ากับระหว่างฟัซซีแอทริบิวส์ อย่างไรก็ตามสำหรับการแก้ไขปัญหาใหม่นั้น เราจะพบว่าการอธิบายปัญหา (Problem description) ที่มีอยู่เหล่านั้นในโลกแห่งความเป็นจริงแล้ว แม้จะเราจะแสดงถึงการอธิบายปัญหาที่เกิดความคลุมเครือได้ แต่ก็มีข้อจำกัดความสามารถในการทำงานโดยเราจำเป็นที่จะต้องดำเนินการเปรียบเทียบด้วยตัวเปรียบเทียบฟัซซีเท่ากับ (FEQ) ระหว่างฟัซซีแอทริบิวส์ มีรายละเอียด คือ

i). พจน์ภาษาฟัซซี (Fuzzy linguistic term) กับพจน์ภาษาฟัซซี (Fuzzy linguistic term)

กรณีตัวอย่างจากเงื่อนไข A FEQ Y THOLD t โดยกรณีของ FEQ (Fuzzy Equal) นั้นจะใช้ข้อมูลค่าดีกรีความคล้ายคลึง (Similarity degree) ในการเปรียบเทียบกันระหว่างค่าพจน์ภาษาฟัซซี *2* ค่าใดๆ ในแอทริบิวต์เดียวกัน ซึ่งวิธีการหาค่าดีกรีความคล้ายคลึง ระหว่าง *2* ฟัซซีเซ็ตใดๆ โดยเราจะได้ผลลัพธ์ความคล้ายคลึงของค่าดีกรีระหว่างคู่พจน์ภาษาฟัซซีที่มีค่ามากกว่าหรือเท่ากับค่า t (Threshold) ตัวอย่างเช่น High FEQ Medium THOLD 0.6 ฉะนั้น เราจะได้ผลลัพธ์ของการค้นหาระหว่างพจน์ฟัซซีทั้งสอง หากเราทำการคำนวณค่าดีกรีความคล้ายคลึง (Similarity degree) ระหว่าง High กับ Medium ที่มากค่ามากกว่าหรือเท่ากับ *0.6* โดยใช้หลักการคำนวณค่าระดับความคล้ายคลึงโดยใช้สูตรทางคณิตศาสตร์ของความสัมพันธ์ระหว่าง Fuzzy Term *กับ* Fuzzy Term ที่มีความกว้างด้านบนไม่ซ้ำซ้อนกัน *(ดังภาพที่ 2.4)* นั้นคือ



ภาพที่ 2.4 แสดงตัวอย่างความคล้ายคลึง (Similarity) ของ *2* พจน์ภาษาฟัซซี 'High' กับ 'Medium'

2.6 การทบทวนวรรณกรรม/สารสนเทศ (Information) ที่เกี่ยวข้อง

ภาวะน้ำหนักเกินและโรคอ้วนจัดเป็นปัญหาสุขภาพที่สำคัญของประชาคมโลก และยังเป็นปรากฏการณ์ที่พบในแทบทุกประเทศ อุบัติการณ์ของภาวะน้ำหนักเกินและโรคอ้วนและความรุนแรงของปัญหาทางสุขภาพที่เกี่ยวข้องมีทิศทางที่เพิ่มขึ้นอย่างรวดเร็ว องค์การอนามัยโลก รายงานว่าใน พ.ศ. 2548 มีประชากรโลกมากถึง 1.6 พันล้านคน จัดอยู่ในภาวะน้ำหนักเกิน และ 400 ล้านคนจัดอยู่ในภาวะอ้วน แต่มีการคาดการณ์ว่าประชากรโลกที่มีภาวะน้ำหนักเกินจะเพิ่มเป็น 2.3 พันล้านคน และโรคอ้วนเพิ่มเป็น 700 ล้านคนใน พ.ศ. 2558 [2] ทิศทางการเปลี่ยนแปลงนี้พบได้ทั้งในประเทศพัฒนาแล้วและกำลังพัฒนา ในแทบทุกกลุ่มประชากรของสังคม และในทุกกลุ่มอายุ องค์การอนามัยโลกรายงานว่า มีประชากรเด็กต่ำกว่า 5 ปี มากถึง 22 ล้านคนอยู่ในภาวะน้ำหนักเกินและอ้วน ในจำนวนนี้ ประมาณสามในสี่อยู่ในประเทศรายได้ปานกลางและต่ำ อย่างไรก็ตาม เราจะพบว่า ในปัจจุบันปริมาณโรคอ้วนของไทยนั้นได้เพิ่มขึ้นอย่างรวดเร็วเช่นกัน และโรคอ้วนนี้เป็นปัญหาที่พบมากขึ้นในเด็กวัยเรียน หากเด็กเหล่านั้นเกิดภาวะของการเป็นโรคอ้วนก็จะทำให้เป็นปัจจัยที่นำไปสู่โรคเรื้อรังได้มากมายเมื่อเติบโตเป็นผู้ใหญ่ สำหรับปัญหาของการเกิดโรคอ้วนในเด็กนั้น โดยส่วนใหญ่แล้วเกิดจากการใช้ชีวิตประจำวันที่มีการบริโภคอาหารเกินความจำเป็นและขาดองค์ความรู้ของผู้ปกครองในการดูแลบุตร รวมทั้งสภาพแวดล้อมทั้งภายในและภายนอกรอบข้างสถานศึกษาของเด็กที่เอื้ออำนวยต่อการเกิดโรคอ้วนขึ้น อย่างไรก็ตามสำนักนโยบายและยุทธศาสตร์ กระทรวงสาธารณสุข ได้รายงานสถิติสถานะสุขภาพคนไทย [4] จากการสำรวจ The Nation Household Education Surveys Program (NHES) ครั้งที่ 3 ปี 2546-2547 เปรียบเทียบกับการสำรวจ NHES ครั้งที่ 4 ปี 2551-2552 พบภาวะไขมันคอเลสเตอรอลในเลือดสูง มีอัตราความชุกเพิ่มขึ้นจากร้อยละ 15.5 เป็น 19.4 กรมอนามัย กระทรวงสาธารณสุข ระบุว่า ผลการสำรวจล่าสุดพบคนไทยมีปัญหาภาวะโภชนาการเกิน และกลายเป็นโรคอ้วนเพิ่มสูงมากที่สุดในรอบ 10 ปี โดยเด็กแรกเกิดไปจนถึงอายุ 12 ปี มีความอ้วนพุ่งสูงถึงร้อยละ 40 รองลงมาคือช่วงอายุ 40-50 ปี อ้วนเพิ่มสูงขึ้นเกือบ 2 เท่า และช่วงอายุ 20-29 ปี มีความอ้วนอยู่ที่ร้อยละ 21.7นอกจากนั้นเราจะพบว่า เด็กไทยที่อ้วนกว่าครึ่งหนึ่ง เมื่อเจาะเลือดหาไขมันในร่างกายแล้ว พบว่าร้อยละ 70 มีปัญหาไขมันสูงเกินมาตรฐานผลสำรวจสุขภาพล่าสุด มีคนไทยอายุ 15 ปี ขึ้นไปเป็นโรคอ้วน ติดอันดับ 5 ของเอเชียแปซิฟิก โดยมีคนอ้วนมากถึง 17 ล้านคนทั่วประเทศ และยังมีแนวโน้มเป็นโรคอ้วนเพิ่มขึ้นอีกประมาณ 4 ล้านคนต่อปี ทำให้รัฐบาลต้องสูญเสียค่าใช้จ่ายในการรักษาพยาบาลมากกว่าปีละ 1 แสนล้านบาท นอกจากนี้เราจะพบว่านักเรียนระดับประถมศึกษาในเขตเมืองกลับมีน้ำหนักเกินเกณฑ์ถึงร้อยละ 15.4 และร้อยละ 13.6 สอดคล้องกับผลการวิจัยของกรมอนามัย พบว่านักเรียนในสังกัดทบวงมหาวิทยาลัยเป็นโรคอ้วนสูงถึงร้อยละ 25.9 - 31.5 รองลงมา เช่น โรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาเอกชน พบเด็กอ้วนร้อยละ 25.7 - 28.1 โรงเรียนในสังกัดสำนักงานการประถมศึกษาแห่งชาติ ร้อยละ 23.3 - 27.4 และโรงเรียนสังกัดกรุงเทพฯ พบอัตราโรคอ้วนในเด็กร้อยละ 11.2 -14.6 และยังพบปัญหาโรคอ้วนในเด็กอนุบาลประมาณร้อยละ 19 ซึ่งในอนาคตเด็กเหล่านี้จะเสี่ยงต่อโรคเรื้อรังต่าง ๆ เช่นโรคเบาหวาน โรคหัวใจ โรคมะเร็งและโรคเรื้อรังต่าง ๆ เพิ่มขึ้น[14] สำหรับสาเหตุสำคัญของโรคอ้วนในคนไทยนั้น คือความนิยมและปริมาณการดื่มเครื่องดื่มประเภทต่างๆ เช่นมีการดื่มน้ำอัดลมมากที่สุด 52.5% เฉลี่ย 3 ขวด (ขวดขนาด 450 มล.) ต่อสัปดาห์ รองลงมาเป็นกาแฟสด 45.2% เฉลี่ย 4.8 แก้ว (ขนาดแก้ว 475 มล.) ต่อสัปดาห์ และอันดับ 3 เป็นนมเปรี้ยว 33.3% เฉลี่ย 3.7 ขวด (ขวดขนาด 400 มล.) เฉลี่ย 3.7 ขวดต่อสัปดาห์ [6] นอกจากนี้ รศ.ดร. วิลาสินี อดุลยานนท์ ผู้อำนวยการสำนักรณรงค์สื่อสารสังคม สสส. ได้กล่าวว่า การปรับเปลี่ยนพฤติกรรมในการบริโภคอาหารเป็นเรื่องยากและต้องใช้เวลา ยิ่งในปัจจุบันนี้ที่มีสิ่งเร้าอยู่รอบตัวเด็กมากไปหมด การสร้างนิสัยและความรู้ให้เด็กได้เลือกหรือตัดสินใจอย่างมีเหตุมีผลต่อสุขภาพที่ดีของตัวเองนั้นเป็นสิ่งสำคัญซึ่ง

จำเป็นต้องใช้การมีส่วนร่วมของผู้บริหารทุกระดับในการกำหนดนโยบายโดยครูและผู้ปกครองมีส่วนสำคัญ ในการให้ความรู้และจัดหาอาหารที่เหมาะสมควบคุมสิ่งแวดล้อมในและรอบโรงเรียนให้ปลอดจากอาหารที่เป็นภัยต่อสุขภาพ [13]

บทที่ 3
วิธีดำเนินการวิจัย

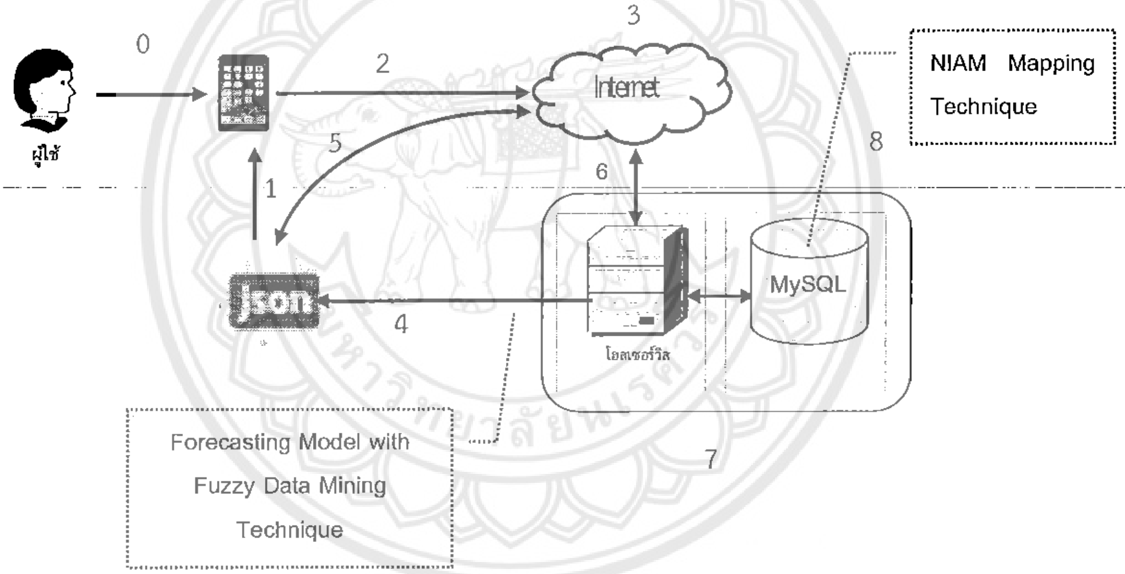
3.1 สถาปัตยกรรมของระบบการทำงานในงานวิจัย

ในการศึกษางานวิจัยนี้ ได้ดำเนินการกำหนดรูปแบบการทำงานโครงสร้างของระบบเพื่อกำหนดขอบเขตตามวัตถุประสงค์ของงานวิจัยได้แบ่งการทำงานออกเป็น 3 ส่วนหลักๆ มีรายละเอียดดังนี้ (ดังภาพที่ 3.1)

ส่วนที่1: เป็นส่วนที่เกี่ยวข้องกับการสร้างโมเดลระบบพยากรณ์จากการเปรียบเทียบระหว่างอัลกอริธึมต้นไม้ตัดสินคลุมเครือกับโครงข่ายประสาทเทียมคลุมเครือภายต้การคัดข้อมูลในลักษณะ cross validation 3 แบบ Percentage split test 3 แบบ และการแยกไฟล์ข้อมูลการทดสอบและข้อมูลการเรียนรู้ 1 แบบ ตามลำดับ นอกจากนี้ได้นำรหัสของโมเดลที่ได้ฝังโค้ดบนเวปแอพพลิเคชันที่พัฒนาภายใต้รูปแบบresponsive (หมายเลข 1, 4, 5)

ส่วนที่2: เป็นส่วนที่ดำเนินการจัดเก็บข้อมูลบนโฮสเซอร์วิส และการออกแบบระบบจัดการฐานข้อมูลภายใต้การทำคอนเซ็บชวลสกีมม่าแมปปิง (Conceptual schema mapping) โดยใช้เทคนิค NIAM (หมายเลข 7, 8)

ส่วนที่3: เป็นส่วนที่เกี่ยวข้องกับผู้ใช้งานผ่าน UI บนอุปกรณ์พกพาเคลื่อนที่ (หมายเลข 0, 2, 3, 6)



ภาพที่ 3.1 แสดงสถาปัตยกรรมของระบบในงานวิจัย

3.2 การรวบรวมปัจจัยที่เกี่ยวข้องของการสร้างแบบจำลองพยากรณ์โรคอ้วน

ปัจจัยที่ส่งผลกระทบต่อการเกิดโรคอ้วนนั้น มีผลงานวิจัยจำนวนมากที่นำเสนอการเกิดโรคอ้วนของช่วงอายุ 15 ปีขึ้นไป โดยที่ใช้หลักการประเมินเกณฑ์การเกิดโรคอ้วนด้วยการคำนวณค่าดัชนีมวลกาย (BMI) ซึ่งแท้จริงแล้วจะนิยมใช้กับผู้ที่อายุเกิน 20 ปี สำหรับงานวิจัยนี้สนใจที่จะใช้ปัจจัยการเกิดโรคอ้วนของเด็กนักเรียนระดับชั้นประถมศึกษาระหว่างอายุ 6-12 ปี จึงไม่จำเป็นที่ใช้ผลการคำนวณค่า BMI เพื่อวัดการเกิดโรคอ้วนเพียงเท่านี้แต่ยังใช้ปัจจัยอื่นๆ ซึ่งเราพบว่าปัจจัยการเกิดโรคอ้วนในประเทศไทยนั้นสาเหตุหลักเกิดจาก 2 ประการ คือ ด้านกรรมพันธุ์ และด้านสิ่งแวดล้อม จากข้อมูลเบื้องต้นเราจะพบว่าปัจจัยด้านกรรมพันธ์

นั้น หากเกิดจากครอบครัวที่มีภาวะโรคอ้วน โดยเฉพาะอย่างยิ่งมารดาเป็นโรคอ้วนย่อมมีผลต่อลูกที่จะเกิดโรคอ้วนสูงมากยิ่งขึ้น และสำหรับปัจจัยสิ่งแวดล้อมนั้นสามารถแบ่งได้เป็น 2 ส่วน นั่นคือ ปัจจัยด้านพฤติกรรมการบริโภคและปัจจัยด้านพฤติกรรมการใช้พลังงาน ตามลำดับ งานวิจัยนี้ใช้ตัวอย่างข้อมูลนำเข้าจำนวน 950 ตัวอย่าง จากนักเรียนระดับชั้นประถมศึกษาของโรงเรียนจำนวนทั้งหมด 125 โรงเรียนในเขตภาคเหนือตอนล่างของประเทศไทย ซึ่งแบ่งเป็นขนาดโรงเรียน 3 ขนาด คือ เล้ก กลางใหญ่ ตามมาตรฐานการแบ่งกลุ่มโรงเรียน โดยกำหนดปัจจัยนำเข้าเริ่มต้นจำนวน 18 แอททริบิวต์ ซึ่งบางแอททริบิวต์ที่เป็นตัวแปรเชิงคุณภาพแบบพหุวิภาค (Polytomous variable) จะมีจำนวน 8 ตัว ดังนั้น เราจะต้องแปลงตัวแปรเชิงคุณภาพให้อยู่ในในรูปแบบที่สามารถวัดค่าได้ นั่นคือ จะต้องสร้างตัวแปรเชิงคุณภาพเหล่านี้ ตัวอย่างเช่น careerParent, educationParent และ class ให้อยู่ในลักษณะตัวแปรหุ่น (Dummy variables) โดยผู้วิจัยได้กำหนดค่าให้เท่ากับ *G*-1 เมื่อ *G* คือจำนวนกลุ่มของตัวแปร ดังตัวอย่าง (ตารางที่3.1) แสดงรายละเอียดตัวแปรหุ่นของระดับชั้นเรียน (class) ตั้งแต่ระดับประถมศึกษาปีที่1ถึงประถมศึกษาชั้นปีที่6 โดยมีการกระทำของตัวแปรระดับชั้นเรียนนี้ให้เป็นตัวแปรทวิภาค (Dichotomous) ชุดใหม่ ที่มีจำนวน G-1 ตัว โดยมีการเลือกกลุ่ม *G6 (นักเรียนระดับชั้นประถมศึกษาปีที่6)* เป็นกลุ่มอ้างอิง (Reference group) เพื่อใช้เป็นกลุ่มเปรียบเทียบสำหรับการดำเนินการจัดเก็บข้อมูลเบื้องต้นมีรายละเอียด *(ตารางที่3.2)* แสดงลักษณะตัวแปรหุ่นของ mealType และมีรายละเอียด (ตารางที่ 3.3) แสดงถึงลักษณะตัวแปรหุ่นของ CongenitalDisease ตามลำดับ โดยได้รวบรวมข้อมูลเบื้องต้นที่นำเสนออยู่ในรูปของไฟล์ testXtaPhu-4.excel ซึ่งมีรายละเอียด (ดังภาพที่ 3.2)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	P	Q	R
1	gender	weight	hight	congenital_	heredity	Free_time_	exercise	fastfood	fried_foods	dessert	beverages	More_ofte	school	class	training	exposition	status
2	Male	39.2	136	unknow	overweight	watch tv a	no	everday	medium	low	sparkling w	much	urban	Grade 3	yes	yes	normal
3	Male	30.2	135	unknow	normal	reading	yes	More than	medium	medium	sparkling w	medium	urban	Grade 3	yes	yes	thin
4	Male	50	138	unknow	fat	eat snacks	no	everday	much	much	sparkling w	very_much	urban	Grade 3	yes	yes	overweight
5	Male	42	134	unknow	overweight	play game	no	More than	medium	medium	sparkling w	much	urban	Grade 3	yes	yes	overweight
6	Male	39.8	145	unknow	normal	exercise	yes	1 per week	medium	medium	green tea	much	urban	Grade 3	yes	yes	thin
7	Male	37.5	138	unknow	normal	reading	yes	everday	medium	medium	sparkling w	medium	urban	Grade 3	yes	yes	thin
8	Male	25.5	129	unknow	normal	play game	yes	everday	medium	much	Chocolate	much	urban	Grade 3	yes	yes	thin
9	Male	37	139	unknow	normal	watch tv a	yes	1 per week	medium	medium	Chocolate	much	urban	Grade 3	yes	yes	thin
10	Male	70	143	unknow	fat	eat snacks	no	everday	much	medium	sparkling w	very_much	urban	Grade 3	yes	yes	fat
11	Male	30	132	unknow	normal	exercise	yes	everday	much	medium	Chocolate	much	urban	Grade 3	yes	yes	thin
12	Male	24	123	unknow	normal	watch tv a	yes	no	low	medium	curd	much	urban	Grade 3	yes	yes	thin
13	Male	30	128	unknow	normal	watch tv a	yes	everday	medium	medium	sparkling w	medium	urban	Grade 3	yes	yes	normal
14	Male	26	125	unknow	fat	reading	no	everday	much	much	sparkling w	much	urban	Grade 3	yes	yes	thin
15	Male	25	134	unknow	normal	reading	no	no	no	low	juice	much	urban	Grade 3	yes	yes	thin
16	Male	37	126	unknow	fat	eat snacks	yes	everday	much	medium	sparkling w	medium	urban	Grade 3	yes	yes	overweight
17	Male	45	137	unknow	overweight	watch tv a	no	everday	medium	medium	Chocolate	much	urban	Grade 3	yes	yes	overweight
18	Male	36	132.5	unknow	overweight	play game	yes	More than	medium	much	Chocolate	very_much	urban	Grade 3	yes	yes	normal
19	Male	42	138	unknow	overweight	play compu	no	More than	much	medium	sparkling w	much	urban	Grade 3	yes	yes	normal
20	Male	50	137	unknow	fat	play game	no	everday	much	much	sparkling w	very_much	urban	Grade 3	yes	yes	overweight
21	Male	42	136	unknow	overweight	watch tv a	yes	everday	medium	medium	Chocolate	medium	urban	Grade 3	yes	yes	normal
22	Female	26	130	unknow	normal	play compu	yes	More than	low	medium	juice	very_much	urban	Grade 3	yes	yes	thin
23	Female	35.3	138	unknow	normal	watch tv a	yes	More than	medium	medium	curd	very_much	urban	Grade 3	yes	yes	normal
24	Female	40	151	unknow	normal	watch tv a	yes	1 per week	medium	medium	curd	much	urban	Grade 3	yes	yes	thin
25	Female	32	130	unknow	normal	play compu	yes	everday	low	medium	curd	much	urban	Grade 3	yes	yes	normal

ภาพที่ 3.2 รายละเอียดข้อมูลเบื้องต้นจากการจัดเก็บเพียงบางส่วนในรูปแบบไฟล์Excel

sex	c2	level6	weight	height	allergy	valvular_t	normal	congenita	exercise	game	money	fast_food	soft_drink	candy	fried	aunt	elementa	no_study	contractor	enterprise	income_p	fat
1	0	0	Medium	Low	0	0	1	0	1	0	1	1	1	1	1	0	1	0	1	0	Medium	Normal
1	0	0	Low	Low	0	0	1	0	1	0	1	1	1	1	1	0	1	0	0	0	High	Normal
1	0	0	Medium	Medium	0	0	1	0	1	0	1	2	2	2	2	0	1	0	1	0	Medium	Normal
1	0	0	High	Low	0	0	1	0	2	0	1	2	2	2	2	0	1	0	1	0	Medium	Fat
1	0	0	Low	Low	0	0	1	0	1	0	1	1	1	1	1	0	1	0	0	0	High	Thin
1	0	0	Low	Low	0	0	1	0	1	0	1	1	1	1	1	0	0	0	0	0	Medium	Thin
1	0	0	Medium	Medium	0	0	1	0	1	0	1	1	1	1	1	0	0	0	1	0	Medium	Normal
1	0	0	Medium	Low	0	0	1	0	2	0	1	2	2	2	2	0	1	0	0	0	High	Normal
1	0	0	Medium	Low	0	0	1	0	1	0	1	1	1	1	1	0	1	0	1	0	High	Normal
1	0	0	Medium	Low	0	0	1	0	2	0	1	2	2	2	2	0	0	0	1	0	High	Normal
1	0	0	Medium	Low	0	0	1	0	1	0	1	1	1	1	1	0	0	0	1	0	Medium	Normal
1	0	0	Low	Low	0	0	1	0	2	0	1	1	1	1	1	0	0	0	1	0	Low	Normal
1	0	0	Medium	Low	0	0	1	0	1	0	1	1	1	2	2	0	0	0	1	0	Medium	Normal
0	0	0	Low	Low	0	0	1	0	2	0	1	1	2	2	2	0	0	0	1	0	Medium	Normal
0	0	0	Low	Low	0	0	1	0	2	0	1	1	1	1	1	0	1	0	1	0	Medium	Thin
0	0	0	Medium	Low	0	0	1	0	1	0	1	2	1	2	1	0	0	0	1	0	High	Normal
0	0	0	Low	Medium	0	0	1	0	2	0	1	2	1	2	2	0	0	0	1	0	Low	Normal
0	0	0	High	Low	0	0	1	0	1	0	1	2	2	3	2	0	1	0	0	0	High	Fat
1	0	0	Medium	Medium	0	0	1	0	1	0	1	0	1	0	1	0	0	0	0	0	High	Normal
1	0	0	Medium	Medium	0	0	0	1	0	0	1	3	0	0	3	0	0	0	0	0	Low	Normal

ภาพที่ 3.3 รายละเอียดการปรับรหัสข้อมูลเพียงบางส่วนในรูปแบบไฟล์ testXtaPhu-4.csv

ตารางที่ 3.1 แสดงตัวอย่าง Dummy variable coding ของตัวแปรระดับชั้นเรียน (Class)

กลุ่ม	Class	X_1	X_2	X_3	X_4	X_5
G1	ประถมศึกษาปีที่1	1	0	0	0	0
G2	ประถมศึกษาปีที่2	0	1	0	0	0
G3	ประถมศึกษาปีที่3	0	0	1	0	0
G4	ประถมศึกษาปีที่4	0	0	0	1	0
G5	ประถมศึกษาปีที่5	0	0	0	0	1
(Reference) G6	ประถมศึกษาปีที่6	0	0	0	0	0

ตารางที่ 3.2 แสดงตัวอย่าง Dummy variable coding ของตัวแปร MealType

กลุ่ม	MealType	X_1	X_2	X_3	X_4
G1	Fast_food	1	0	0	0
G2	Soft_drink	0	1	0	0
G3	Candy	0	0	1	0
G4	Fried	0	0	0	1
(Reference) G5	Late_meal	0	0	0	0

ตารางที่ 3.3 แสดงตัวอย่าง Dummy variable coding ของตัวแปร *CongenitalDisease*

กลุ่ม	Congenital Disease	X_1	X_2	X_3	X_4	X_5	X_6
G1	Asthma	1	0	0	0	0	0
G2	Allergy	0	1	0	0	0	0
G3	Thalassemia	0	0	1	0	0	0
G4	Valvular_heart	0	0	0	1	0	0
G5	Mumps	0	0	0	0	1	0
G6	Gastritis	0	0	0	0	0	1
(Reference) G7	Normal	0	0	0	0	0	0

3.3 การคัดเลือกปัจจัยนำเข้าด้วยการวิเคราะห์สมการสหสัมพันธ์เพียร์สัน (Pearson's correlation analysis)

งานวิจัยนี้ได้นำหลักการวิเคราะห์สหสัมพันธ์ระหว่างตัวแปรเพื่อทำการวัดระดับค่าตัวแปรในมาตราอันตรภาค ภายใต้รูปแบบการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์เพียรสัน (Pearson correlation coefficient) หรือบางครั้งเรียกว่า สหสัมพันธ์อย่างง่าย *(*Simple Correlation*)* สหสัมพันธ์เชิงเดี่ยว ซึ่งเป็นการหาความสัมพันธ์ระหว่างตัวแปร *2* ตัว *(*Bivariate Correlation*)* โดยที่เราเรียกว่าตัวแปรอิสระว่า ตัว

ตารางที่ 3.4 แสดงรายละเอียดความสัมพันธ์ระหว่างตัวแปรพยากรณ์กับตัวแปรตามด้วยค่าเพียร์สัน

weight	height	Level6	Fast_food	congenital
0.864	0.824	0.712	0.711	0.689
Soft_drink	fried	game	money	C2
0.667	0.659	0.612	0.604	0.586
Income_parent	normal	exercise	contractors	aunt
0.55	0.543	0.540	0.532	0.518
Enterprise_ employee	candy	sex	Valvular_heart	Elementary_ education
0.479	0.472	0.469	0.458	0.455
No_study	allergy			
0.442	0437			

แปรทำนาย *(Predictor variable)* และเรียกตัวแปรอีกตัวว่าตัวแปรเกณฑ์ (Criterion variable) ซึ่งปกติจะเป็นตัวแปรตาม ดังนั้น งานวิจัยนี้ได้คัดเลือกปัจจัยนำเข้าภายใต้การใช้ขนาดตัวอย่างที่มีปริมาณมากกว่า 20เท่าของตัวแปรอิสระ และได้กำหนดให้ Fat เป็นตัวแปรตาม สำหรับการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์เพียรสัน (Pearson's Correlation Coefficient) โดยใช้สัญลักษณ์ r หรือ r_{XY} (ดังสมการ 1) เพื่อคำนวณค่า r ซึ่งในงานวิจัยนี้ได้คัดเลือกตัวแปรอิสระที่แปลงอยู่ในรูปแบบตัวแปรเชิงปริมาณแล้วโดยคัดเลือกจากค่า $0.5 \leq r \leq 1.0$ จำนวนทั้งสิ้น 22 ตัวแปรอิสระ ดังตัวอย่างของตัวแปรทำนายที่มีค่าสูงสุด *5* ลำดับแรกได้แก่ ตัวแปร weight, height, level6, fast_food และ congenital ซึ่งมีค่าเท่ากับ 0.864, 0.824, 0.712, 0.711 และ 0.689 ตามลำดับมีรายละเอียด (ตารางที่ 3.4)

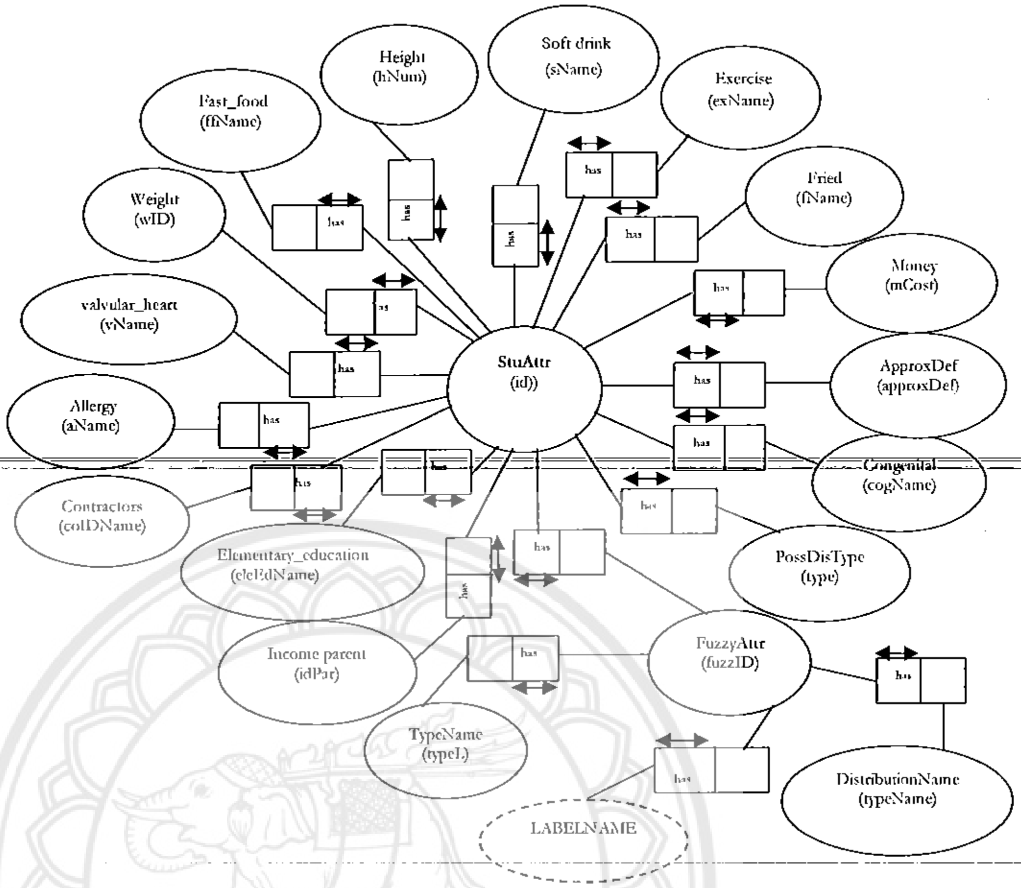
$$r_{XY} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}} \qquad (1)$$

เมื่อ r_{XY} = แทนค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

$\sum X$ = แทนผลรวมของข้อมูลทั้งหมดในตัวแปร X

$\sum Y$ = แทนผลรวมของข้อมูลทั้งหมดในตัวแปร Y

N = แทนจำนวนข้อมูลในแต่ละตัวแปร



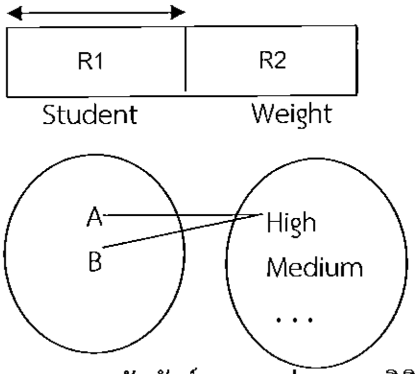
ภาพที่ 3.5 จำลองความสัมพันธ์ของข้อมูลในระดับแนวคิดด้วย NIAM Conceptual Schema

3.4 การออกแบบระบบจัดการฐานข้อมูล

3.4.1 การออกแบบความสัมพันธ์ของข้อมูลในระดับเชิงแนวคิด

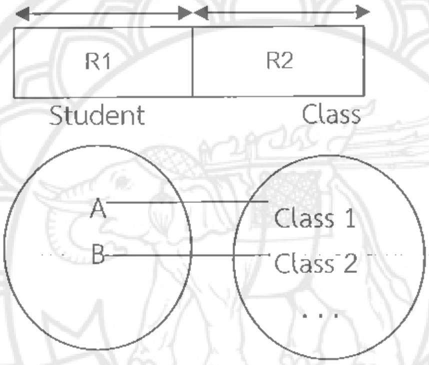
ในงานวิจัยได้นำวิธีการทางไนแอม (NIAM: Nijssen's Information System Analysis Method) ซึ่งไนแอม (NIAM) เป็นการนำเสนอรูปแบบสร้างความสัมพันธ์ของข้อมูลในระดับแนวคิด ซึ่งจะประกอบไปด้วยชนิดเอนติตี้ (Entity type) ชนิดเลเบิ้ล (Label type)ชนิดความจริง (Fact type)ชนิดอ้างอิง (Reference type)และข้อจำกัดเพื่อความถูกต้องของข้อมูล (Integrity constraints) เราสามารถออกแบบความสัมพันธ์อย่างง่ายๆ (ดังภาพที่ 3.5) โดยรูปแบบการออกแบบความสัมพันธ์ระดับแนวคิดในงานวิจัยนี้สามารถแสดงได้ดังตัวอย่างบางส่วนต่อไปนี้คือ (ดังภาพที่ 3.6 ภาพที่3.7 และ ภาพที่ 3.8 ตามลำดับ)

รูปแบบ Many-to-One แสดงรายละเอียดของแอททริบิวต์ Student มีความสัมพันธ์กับแอททริบิวต์ Weight แบบ Many to one นั่นคือ นักเรียนจำนวนหลายๆ คน สามารถมีน้ำหนักในรูปแบบ High หรือกล่าวอีกอย่างก็คือ น้ำหนัก High จะเกิดขึ้นกับนักเรียนจำนวนหลายๆ คน



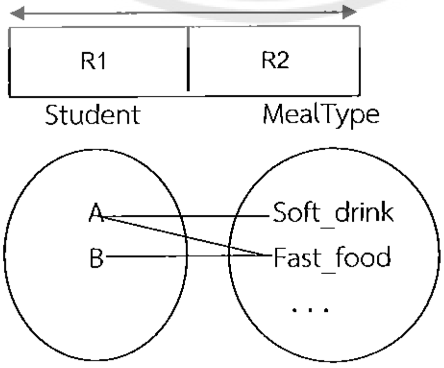
ภาพที่ 3.6 ความสัมพันธ์ของระหว่างแอททริบิวต์แบบ Many-to-One

รูปแบบ One-to-One แสดงรายละเอียดของแอททริบิวต์ Student มีความสัมพันธ์กับแอททริบิวต์ Class แบบ One to one นั่นคือ นักเรียนแต่ละคน จะอยู่ได้เพียงระดับชั้นเรียนเดียวเท่านั้น



ภาพที่ 3.7 ความสัมพันธ์ของระหว่างแอททริบิวต์แบบ Many-to-One

รูปแบบ Many-to-Many แสดงรายละเอียดของแอททริบิวต์ Student มีความสัมพันธ์กับแอททริบิวต์ MealType แบบ Many to Many นั่นคือ นักเรียน 1 คน สามารถทานอาหารได้หลายรูปแบบ หรือกล่าวอีกอย่างก็คือ รูปแบบการทานอาหาร 1 รูปแบบ จะเกิดขึ้นกับนักเรียนจำนวนหลายๆ คน

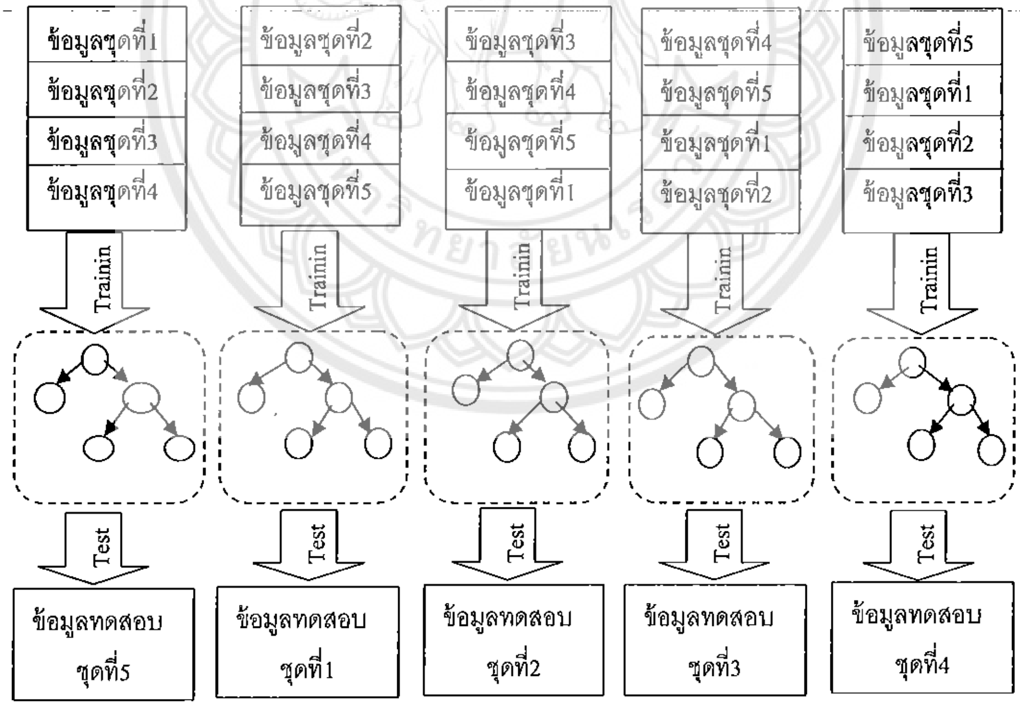


ภาพที่ 3.8 ความสัมพันธ์ของระหว่างแอททริบิวต์แบบ Many-to-One

3.5 การสร้างแบบจำลองพยากรณ์โรคอ้วนด้วยเทคนิคฟัซซีนิวรอลเน็ตเวิร์คกับ ฟัซซีต้นไม้ช่วยตัดสินใจ

ผู้วิจัยได้ดำเนินการเพื่อเลือกรูปแบบการแบ่งข้อมูลเพื่อใช้วัดประสิทธิภาพของตัวแบบจำลอง โดยผู้วิจัยได้เลือกการแบ่งข้อมูล 2 รูปแบบหลักๆ นั่นคือรูปแบบแรกจะใช้ลักษณะ Split Test (ใช้ข้อมูลตัวอย่างฝึกสอน 34% : 70% และ 80% โดย ใช้ข้อมูลตัวอย่างทดสอบ 66%: 30% และ80% ตามลำดับ และรูปแบบที่สองคือวิธี Cross-validation Test ที่ผู้วิจัยจะกำหนดค่า K=5 และ K=10 เท่านั้น เพราะมีความเหมาะสมในจำนวนครั้งที่ทำการทดสอบและเป็นที่นิยม เพื่อใช้ในการวัดประสิทธิภาพของโมเดล (ดังภาพที่ 3.9 ภาพที่ 3.10 ภาพที่ 3.11 และภาพที่ 3.12 ตามลำดับ)

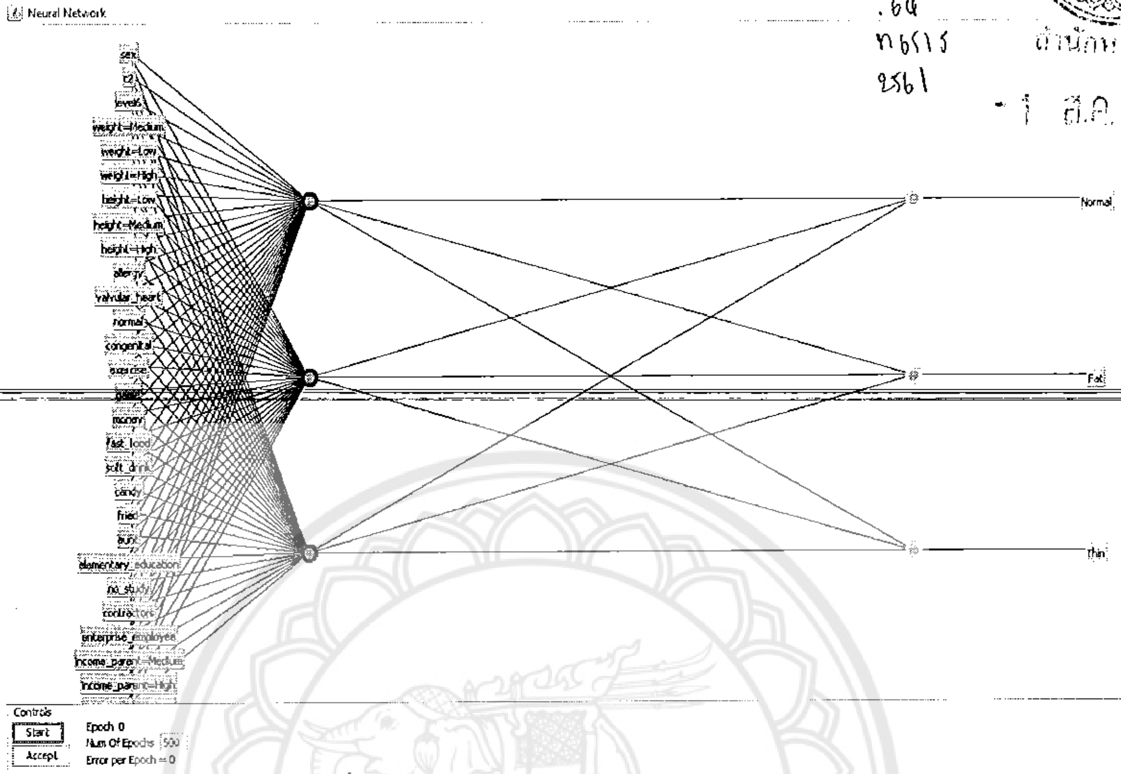
- ในรายงานฉบับนี้ ผู้วิจัยขอยกตัวอย่างหลักการแบ่งข้อมูลของผู้วิจัย กรณีการแบ่งข้อมูลในลักษณะ 5-fold cross-validation ซึ่งกำหนดให้ค่า K=5 ด้วยการแบ่งข้อมูลออกเป็น 5 ส่วน จากจำนวนข้อมูลตัวอย่างที่นำเสนอในรายงานนี้ 965 ตัวอย่างมีความเหมาะสมเมื่อแบ่งเป็นส่วนๆ จะได้ส่วนละ 193 ตัวอย่าง โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน
- หลักการวนรอบของการทำงานจากข้อมูลตัวอย่างเพื่อนำมาสร้างโมเดลและใช้เป็นโมเดลทำนายเพื่อหาประสิทธิภาพที่ดีที่สุดของโมเดล นั่นก็คือ ผู้วิจัยกำลังดำเนินการดังนี้
  - รอบที่ 1 ใช้ข้อมูลส่วนที่ 1,2,3 และ 4 สร้างโมเดล และใช้โมเดลทำนายข้อมูลส่วนที่ 5
  - รอบที่ 2 ใช้ข้อมูลส่วนที่ 2,3,4 และ 5 สร้างโมเดล และใช้โมเดลทำนายข้อมูลส่วนที่ 1
  - รอบที่ 3 ใช้ข้อมูลส่วนที่ 3,4,5 และ 1 สร้างโมเดล และใช้โมเดลทำนายข้อมูลส่วนที่ 2
  - รอบที่ 4 ใช้ข้อมูลส่วนที่ 4,5,1 และ 2 สร้างโมเดล และใช้โมเดลทำนายข้อมูลส่วนที่ 3
  - รอบที่ 5 ใช้ข้อมูลส่วนที่ 5,1,2 และ 3 สร้างโมเดล และใช้โมเดลทำนายข้อมูลส่วนที่ 4



ภาพที่ 3.9 แสดงการแบ่งข้อมูลตัวอย่างใช้ทดสอบประสิทธิภาพตัวแบบจำลองลักษณะ5 fold cross validation

1020391
ว QA
9
.64
ท6515
2561
สำนักหอสมุด
- 1 ส.ค. 2562

- **3.5.1 การสร้างแบบจำลองด้วยเทคนิคฟัซซีนิวรอลเน็ตเวิร์ค**



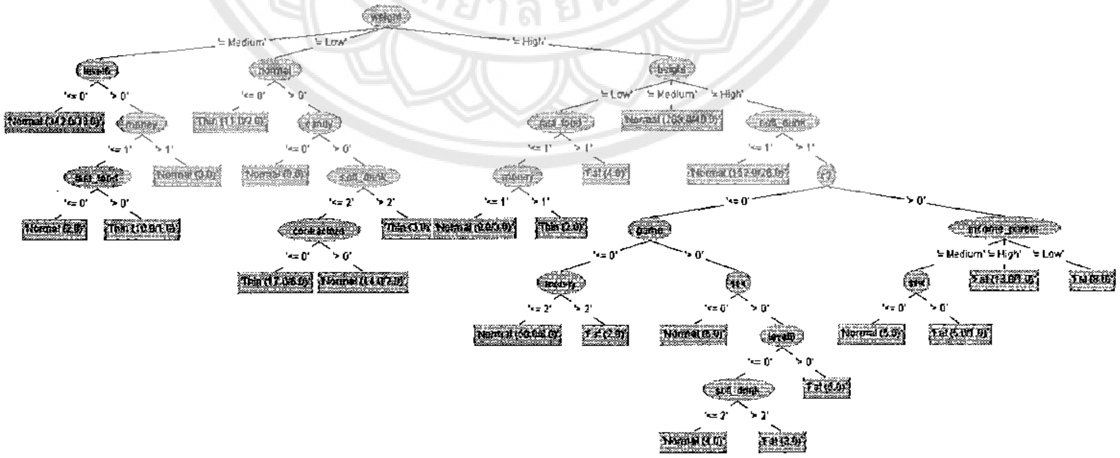
ภาพที่ 3.10 โครงสร้างนิวรอลเน็ตเวิร์ครูปแบบ 28-3-3

```
=== Classifier model (full training set) ===

Sigmoid Node 0
    Inputs    Weights
    Threshold    -10.832468958339723
    Node 3    4.017478465283201
    Node 4    6.433859724115343
    Node 5    8.509341942994585
Sigmoid Node 1
    Inputs    Weights
    Threshold    7.731846011613798
    Node 3    -4.805872294596979
    Node 4    -12.028924653545305
    Node 5    -4.635878700766036
Sigmoid Node 2
    Inputs    Weights
    Threshold    -4.024965674337025
    Node 3    -0.14316725098210076
    Node 4    4.818472595994004
    Node 5    -3.5423146251794155

    . . .
```

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         775               84.3109 %
Incorrectly Classified Instances       190               15.6891 %
Kappa statistic                          0.2505
Mean absolute error                      0.1649
Root mean squared error                  0.3296
Relative absolute error                 75.5629 %
Root relative squared error             99.949  %
Total Number of Instances              965

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.933     0.75      0.841     0.933     0.885      0.679    Normal
                 0.275     0.037     0.484     0.275     0.351      0.789    Fat
                 0.213     0.022     0.444     0.213     0.288      0.693    Thin
Weighted Avg.    0.843     0.613     0.827     0.843     0.828      0.693

=== Confusion Matrix ===

   a   b   c   <-- classified as
 729  32  20 |   a = Normal
  79  30   0 |   b = Fat
  59   0  16 |   c = Thin
```

ภาพที่ 3.11 แสดงรายละเอียดโหนดบางส่วนของการใช้เทคนิคนิวรอลเน็ตเวิร์ค Cross-validation fold(10)

- 3.5.2 การสร้างแบบจำลองด้วยเทคนิคฟัซซีต้นไม้ช่วยตัดสินใจ



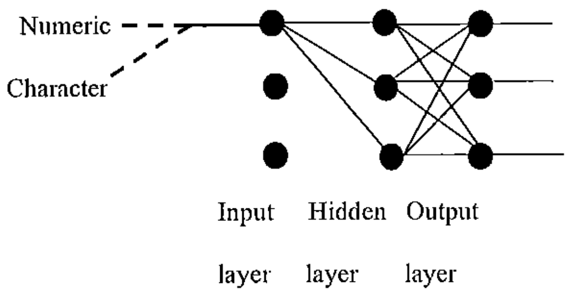
```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          812               84.1451 %
Incorrectly Classified Instances        153               15.8549 %
Kappa statistic                           0.3557
Mean absolute error                       0.1645
Root mean squared error                   0.3023
Relative absolute error                  75.3918 %
Root relative squared error              91.6672 %
Total Number of Instances               965

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.972     0.712     0.853     0.972     0.908      0.659    Normal
                 0.294     0.009     0.8       0.294     0.43       0.775    Fat
                 0.28      0.016     0.6       0.28      0.382      0.745    Thin
Weighted Avg.    0.841     0.578     0.827     0.841     0.813      0.679

=== Confusion Matrix ===

   a   b   c   <-- classified as
 759   8  14 |   a = Normal
  77  32   0 |   b = Fat
  54   0  21 |   c = Thin
```

ภาพที่ 3.12 แสดงรายละเอียดโหนดบางส่วนของการใช้เทคนิคต้นไม้ช่วยตัดสินใจ Cross-validation fold(10)

3.6 การนำเสนอ Fuzzy data mining technique

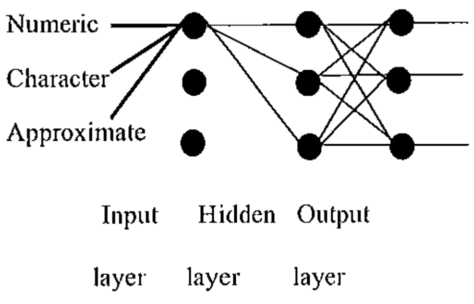
การประยุกต์ใช้ข้อมูลในลักษณะเชิงฟัซซี เป็นตัวกำหนดให้สอดคล้องกับการนำเข้าของข้อมูลที่แท้จริง เพื่อให้การทำงานของอัลกอริธึมมีความเหมาะสมมากยิ่งขึ้น ทั้งนี้ปกติแล้วการทำงานของหลักการอัลกอริธึม เช่น โครงข่ายประสาทเทียมนั้น จะประกอบไปด้วยหลักการทำงาน 3 ส่วน นั่นคือ ส่วนชั้นข้อมูลนำเข้า (Input layer) ส่วนชั้นซ่อน (Hidden layer) และส่วนของผลลัพธ์ (Output layer) อย่างไรก็ตามในงานวิจัยนี้ได้นำเสนอรูปแบบของข้อมูลนำเข้าในลักษณะเชิงฟัซซี ทำให้การใช้งานของข้อมูลจะมีลักษณะดังต่อไปนี้ (ดังภาพที่ 3.13 และ ภาพที่ 3.14 ตามลำดับ)

รูปแบบ การนำเข้าข้อมูลชนิดเดียว



ภาพที่ 3.13 Neural net

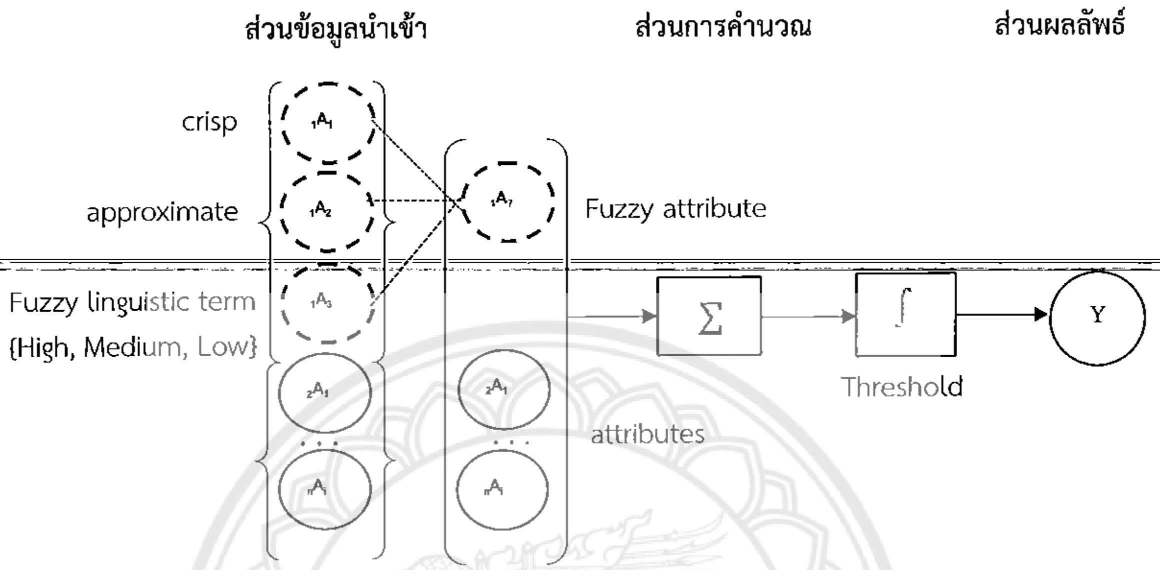
รูปแบบการนำเขาข้อมูลหลายๆ ชนิด



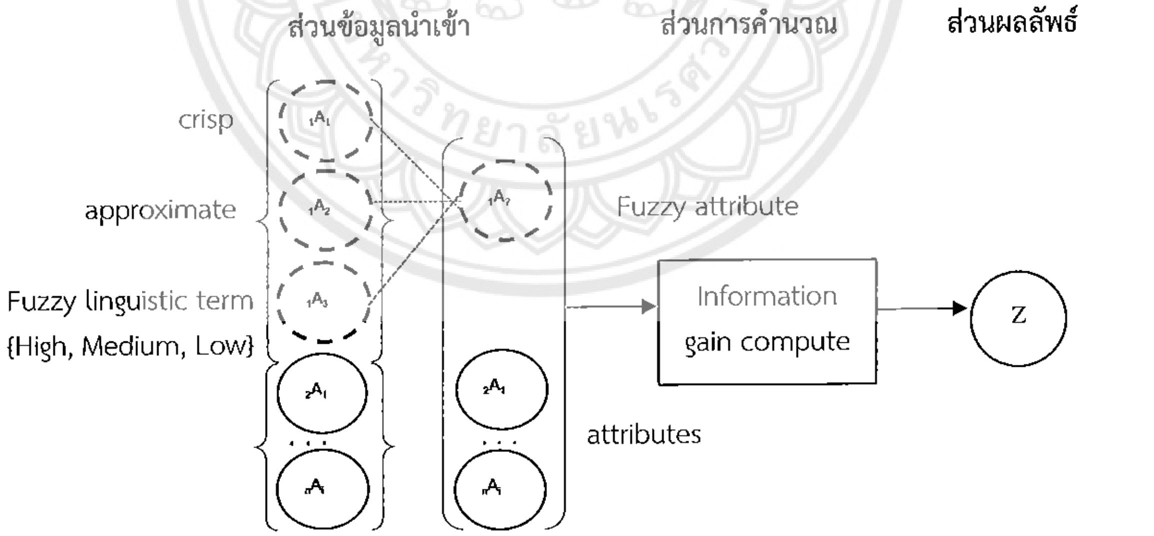
ภาพที่ 3.14 Fuzzy neural net

3.7 หลักการทำงานของ Fuzzy neural network และ Fuzzy decision trees

ในงานวิจัยนี้ ได้ดำเนินการภายใต้การกำหนดให้ฟัซซีแอททริบิวต์ของน้ำหนัก ส่วนสูง และค่ารายได้ของครอบครัวกรณีที่รูปในลักษณะพจน์ภาษาฟัซซี จะมีระดับของสมาชิกคือ High Medium Low ตามลำดับ ซึ่งทำให้แอททริบิวต์ดังกล่าว



ภาพที่ 3.15 แสดงการทำงานของ Fuzzy neural network model



ภาพที่ 3.16 แสดงการทำงานของ Fuzzy decision trees model

มีการเพิ่มจำนวนโหนดนำเข้า อย่างละ 3 โหนด เพื่อเป็นปัจจัยร่วมสำหรับใช้สร้างตัวแบบจำลองพยากรณ์ ดังนั้น ในงานวิจัยนี้ตัวแบบจำลองพยากรณ์สามารถรองรับการนำเข้าข้อมูลเพื่อทดสอบการพยากรณ์ภายใต้

ลักษณะข้อมูลนำเข้าได้ทั้ง 3 รูปแบบ นั่นคือ ค่าคริสป์ ค่าประมาณ และค่าพจน์ภาษาฟัซซี ได้ตามลำดับ ภายใต้แนวคิดของการแมทชิ่งค่าของฟัซซีแอททริบิวต์แล้วเลือกเป็นโหนดตัวแทนเพื่อใช้ดำเนินการคำนวณของการสร้างผลลัพธ์ของการพยากรณ์ ซึ่งรายละเอียดการทำงานของตัวแบบจำลองที่จะต้องดำเนินการคัดกรองการนำเข้าข้อมูลดังที่กล่าวข้างต้นนี้ สามารถแสดงรายละเอียดทั้งรูปแบบของ Fuzzy neural network model และรูปแบบของ Fuzzy decision tree model ตามลำดับ ซึ่งทั้งสองอัลกอริธึมนี้ ได้กำหนดให้การดำเนินการส่วนของข้อมูลนำเข้าภายใต้แนวคิดเดียวกัน ดังมีรายละเอียด (ดังภาพที่ 3.15 และ ภาพที่ 3.16) ต่อไปนี้

3.8 การนำเสนอค่าฟัซซีแอททริบิวต์ (Fuzzy attribute value representation)

การดำเนินการเพื่อจำแนกข้อมูล (Data classification) ภายใต้เทคนิคการทำเหมืองข้อมูล ซึ่งในงานวิจัยนี้ได้นำเสนอเพื่อเปรียบเทียบการใช้เทคนิคต้นไม้ตัดสินใจ (Decision tree) และโครงข่ายประสาทเทียม (Neural net)ที่ทำงานในลักษณะเชิงฟัซซี สร้างตัวแบบพยากรณ์ปริมาณโรคอ้วน ตามลำดับ โดยก่อนดำเนินการเข้าสู่การใช้อัลกอริธึมต้นไม้ช่วยตัดสินใจและโครงข่ายประสาทเทียมนั้น ในงานวิจัยจะดำเนินการปรับเปลี่ยนข้อมูลสัดส่วนตามมาตรฐานของนักเรียนของแต่ละบุคคลเพื่อเป็นข้อมูลระดับของแต่ละโรงเรียน เพราะฉะนั้นการนำเข้าข้อมูลมีโอกาสที่จะมีลักษณะเชิงฟัซซี เนื่องจากเป็นการตอบคำถามค่าที่ได้อาจจะเป็นค่าคาดการณ์เท่านั้นของตัวนักเรียนที่ถูกสุ่ม ซึ่งในงานวิจัยทั่วๆ ไปนั้น เป็นเพียงการใช้ข้อมูลที่มีค่าข้อมูลชนิดเดียวกันภายใต้การทำงานของค่าแอททริบิวต์ ดังเช่น น้ำหนักของนักเรียน (weight) มีการกำหนดเป็นตัวเลข เช่น น้ำหนัก 30 กิโลกรัม หรือน้ำหนัก 25 กิโลกรัม เป็นต้น ซึ่งในความเป็นจริงแล้ว ค่าข้อมูลสามารถแสดงได้หลากหลายกว่านี้ ซึ่งจากงานวิจัยนี้ได้นำเสนอการรูปแบบของข้อมูลที่ใช้งานใน 3 ลักษณะ (ดังตารางที่ 3.5) แม้จะเป็นแอททริบิวต์เดียวกัน นั่นคือ ตัวอย่างเช่น แอททริบิวต์ของน้ำหนักนักเรียน ซึ่งจะแสดงได้คือ รูปแบบที่1 คือ ค่าพจน์ภาษาฟัซซีได้แก่ high medium low ซึ่งจะสอดคล้องกับค่าที่กำหนดไว้ในรูปแบบสี่เหลี่ยมคางหมูที่มีการกำหนดขอบเขตของข้อมูลภายใต้ตัวแปร $\propto, \beta, \gamma$ และ δ ตามลำดับ ดังตัวอย่างข้อมูล *weight* และส่วนสูง ที่ปรากฏ *(ดังตารางที่ 3.6 และตารางที่3.7 ตามลำดับ)* สำหรับรูปแบบที่2 คือ ค่าคริสป์ เช่น น้ำหนัก 35 กิโลกรัม และรูปแบบที่3 คือ ค่าประมาณ เช่น ประมาณ 25 กิโลกรัม เป็นต้น

ตารางที่ 3.5 แสดงค่าระหว่างแอททริบิวต์ (Attributes)กับฟัซซีแอททริบวต์ (Fuzzy attributes)

Attribute (Decision tree/ Neural net)		**Fuzzy attribute** (Decision tree/ Neural net)	
Weight(Kg.)	Height(cm.)	Weight(Kg.)	Height(cm.)
30	128	30	Approx128
25	130	Medium	130
38	137	Approx38	High

ตารางที่ 3.6 แสดงค่าลักษณะพจน์ภาษาเชิงฟัซซีของแอททริบิวต์ $attr_{11}$: *"น้ำหนักของนักเรียน" (หน่วย: กิโลกรัม)*

fuzzy ID	label name	attrName	type	DistributionName			
				α	β	γ	δ
0011	weight	High	Trapezoidal	30	55	75	98
		Medium	Trapezoidal	21	26	31	37
		Low	Trapezoidal	10	12	18	23

ตารางที่ 3.7 แสดงค่าลักษณะพจน์ภาษาเชิงฟัซซีของแอททริบิวต์ $attr_{12}$: *"ส่วนสูงของนักเรียน (หน่วย: เซ็นติเมตร)"*

fuzzy ID	label name	attrName	type	DistributionName			
				α	β	γ	δ
0012	height	High	Trapezoidal	140	148	155	165
		Medium	Trapezoidal	120	130	135	142
		Low	Trapezoidal	102	110	118	125

ตารางที่ *3.8* แสดงค่าของคำตอบภายใต้ลักษณะพจน์ภาษาเชิงฟัซซีของแอททริบิวต์ $attr_{11}$: **"รายได้ผู้ปกครองของนักเรียน"**

fuzzyID	labelname	attrName	type	DistributionName($x_{j:1..4}$)			
				α	β	γ	δ
0013	Income_parent	High	Trapezoidal	35K	60K	75K	98K
		Medium	Trapezoidal	21K	26K	31K	37K
		Low	Trapezoidal	10K	12K	18K	23K

3.9 การทำงานของ Fuzzy Attribute Matching Techniques

แนวคิดของการดำเนินการแมทชิ่งค่าแอททริบิวต์ภายใต้การทำงานของแบบจำลองที่เหมาะสมนั้น ในงานวิจัยนี้ได้นำเสนอหลักการแมทชิ่งข้อมูลนำเข้าที่มีลักษณะเป็นฟัซซีแอททริบิวต์ที่ประกอบไปด้วย ค่าน้ำหนัก (Weight) ค่าส่วนสูง (Height) และค่ารายได้ของครอบครัว (Income_parent) โดยข้อมูลนำเข้านั้นสามารถเป็นค่าพจน์ภาษาฟัซซี หรือค่าประมาณ หรือค่าคริสป์ ได้ตามลำดับ ดังนั้น ลักษณะของการแมทชิ่งที่เกิดขึ้นภายใต้งานวิจัยนี้จะดำเนินการภายใต้รายละเอียดดังนี้

รูปแบบที่ 1. การทำ Fuzzy attribute matching ในลักษณะ ค่าคริสป์ (Crisp) กับพจน์ภาษาฟัซซี (Fuzzy linguistic term) ซึ่งเป็นหาค่าจากการเปรียบเทียบค่าคริสป์กับพจน์ภาษาฟัซซี ว่าอยู่ในระดับดีกรีความเป็นสมาชิกที่ดีกรีเท่าใด

$$\mu(x:\alpha,\beta,\gamma,\delta) = \begin{cases} 1 \; when \; \beta \le x \le \gamma \\ 0 \; when \; x \le \alpha \; or \; x \ge \delta \\ \frac{\alpha-x}{\alpha-\beta} \; or \; \alpha < x < \beta \\ \frac{\delta-x}{\delta-\gamma} \; when \; \gamma < x < \delta \end{cases}$$

โดยที่ x คือค่าคริสป์ใดๆ

รูปแบบที่ 2. การทำ Fuzzy attribute matchingในลักษณะ ค่าประมาณ (Approximate value) กับ พจน์ภาษาฟัซซี (Fuzzy linguistic term) โดยค่าประมาณ (Approx x) จะนิยามโดยใช้ฟังก์ชันสมาชิกรูปสามเหลี่ยม (Triangle membership function) โดยการเปรียบเทียบกับค่าพจน์ภาษาฟัซซีหาได้จากความสูงของจุดของส่วนซ้อนทับ (Intersection) ที่สูงที่สุด 2 ฟังก์ชันสมาชิกใดๆ โดยเงื่อนไขของสมการเป็นดังต่อไปนี้

$$\mu(x,a:\alpha,\beta,\gamma,\delta) = \begin{cases} 1 \; when \; \beta \le x \le \gamma \\ 0 \; when \; x + a \le \alpha \; \text{or} \; x - a \ge \delta \\ \frac{\delta-x+a}{\delta-\gamma+a} \; when \; x > \gamma \; and \; x - a < \delta \\ \frac{x+a-\alpha}{a+\beta-\alpha} \; when \; x < \beta \; \; and \; x + a > \alpha \\ max(k_1, k_2); k_1 = \frac{x+a-\alpha}{(\beta-\alpha)+a}, k_2 = \frac{\alpha-x+a}{(\beta-\alpha)-a} \\ \quad when \; (x - a \ge \alpha \; \; and \; \alpha x < \beta) \\ \quad or \; k_1 = \frac{\delta-x+a}{(\delta-\gamma)+a}, k_2 = \frac{\delta-x-a}{(\delta-\gamma)-a} \\ \quad or \; (x + a \le \delta \; and \; \gamma < x < \delta) \end{cases}$$

โดยที่ x is approximate value

a is margin value of graph

ฟังก์ชันสมาชิก รูปสามเหลี่ยม

รูปแบบที่ 3. การทำ Fuzzy attribute matchingในลักษณะค่า พจน์ภาษาฟัซซี (Fuzzy linguistic term) กับ พจน์ภาษาฟัซซี (Fuzzy linguistic term จะทำการเปรียบเทียบกันระหว่างค่าพจน์ภาษาฟัซซี *2* ค่าใดๆ ในแอทริบิวต์เดียวกัน ซึ่งวิธีการหาค่าดีกรีความคล้ายคลึง ระหว่าง *2* ฟัซซีเซ็ตใดๆ โดยเราจะได้ผลลัพธ์ความคล้ายคลึงของค่าดีกรีระหว่างคู่พจน์ภาษาฟัซซีที่มีค่ามากกว่าหรือเท่ากับค่า t (Threshold)

$$S = k/(\frac{w_1-w_2}{\delta_1-\alpha_2} - k) \quad \text{โดยที่ } k = \frac{\delta_1-\alpha_2}{\delta_1-\gamma_1+\delta_2-\gamma_2}$$

$$w_i = w_{bi} + w_{ti} \; ; i = 1,2$$

ความกว้างด้านล่าง (bottom width) $w_{bi} = \frac{(\delta_i-\alpha_i)}{2}$

ความกว้างด้านบน (top width) $w_{ti} = \frac{(\gamma_i-\beta_i)}{2}$

บทที่4
ผลการทดลอง

การวัดประสิทธิภาพของตัวแบบจำลองพยากรณ์นั้น สามารถอธิบายได้นั่นคือ ผลจากการทดสอบจะพบว่า อัลกอริธึมของต้นไม้ช่วยตัดสินใจที่ใช้อัลกอริธึม J48 ภายใต้การแบ่งข้อมูลทดสอบแบบ Cross-validation folds = 10 มีค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล มีค่าเท่ากับ 84.1%, 82.7%, 84.1% และ 81.3% ตามลำดับ สำหรับอัลกอริธึมของต้นไม้ช่วยตัดสินใจที่ใช้อัลกอริธึม ID3 ภายใต้การแบ่งข้อมูลทดสอบแบบ Cross-validation folds =10 มีค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล มีค่าเท่ากับ 79.5%, 80.1%, 79.5% และ 79.7% ตามลำดับ สำหรับอัลกอริธึมโครงข่ายประสาทเทียม ที่มีจำนวนโหนดซ่อน 3 โหนด โมเมนตัม 0.2 อัตราการเรียนรู้ 0.3 ภายใต้การแบ่งข้อมูลทดสอบแบบ Cross-validation folds = 10 มีค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล มีค่าเท่ากับ 84.3%, 82.7%, 84.3% และ 82.8% ตามลำดับ (ดังตารางที่ 4.1) ดังนั้น งานวิจัยนี้ จะพบว่าตัวแบบจำลองพยากรณ์โรคอ้วนนั้น จะใช้แบบจำลองที่มีโครงสร้างแบบโครงข่ายประสาทเทียม 28-3-3 โดยมีรายละเอียดบางส่วนของโครงสร้างดังนี้

ตารางที่ 4.1 แสดงรายละเอียดประสิทธิภาพของอัลกอริธึมที่ทดสอบ

Classifier		Test options	Correctly	Precision	Recall	F-Measure
TREES	ID3	Percentage split20%	74.4	74.1	74.4	74.2
		Percentage split66%	75.2	73.8	75.2	74.4
		Cross validation 5-folds	79.0	79.5	79.0	79.2
		Cross validation 10-folds	79.5	80.1	79.5	79.7
	J48	Percentage split20%	81.5	71.0	81.5	75.2
		Percentage split66%	79.6	77.8	79.6	77.9
		Cross validation 5-folds	82.6	80.0	82.6	80.1
		Cross validation 10-folds	84.1	82.7	84.1	81.3
ANN	HN3	Percentage split20%	71.0	72.7	71.0	71.7
		Percentage split66%	78.4	75.0	78.4	75.9
		Cross validation 5-folds	82.8	81.9	82.8	81.6
		Cross validation 10-folds	84.3	82.7	84.3	82.8
	HN5	Percentage split20%	74.9	73.0	74.9	73.9
		Percentage split66%	79.0	77.2	79.0	77.8
		Cross validation 5-folds	80.1	77.7	80.1	78.6
		Cross validation 10-folds	83.2	82.6	83.2	82.4

ในการดำเนินการวิจัยนี้ พบว่าแบบจำลองตัวพยากรณ์โรคอ้วนที่เหมาะสมนั้นคือ อัลกอริธึมโครงข่ายประสาทเทียมมีโครงสร้างแบบโครงข่ายประสาทเทียม 31-3-3 โมเมนตัม 0.2 อัตราการเรียนรู้ 0.3 ภายใต้การแบ่ง

ข้อมูลทดสอบแบบ Cross-validation folds = 10 มีค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล มีค่าเท่ากับ 84.3%, 82.7%, 84.3% และ 82.8% ตามลำดับ ซึ่งมีค่าน้ำหนัก (weight value) ระหว่าง Input node กับ Hidden node และระหว่าง Hidden node กับ Output node (ดังตารางที่ 4.2 และ ตารางที่4.3) ตามลำดับ

ตารางที่ 4.2 แสดงตารางข้อมูลบางส่วนของน้ำหนักระหว่าง Input node และ Hidden node

List of input node	Sigmoid Node3	Sigmoid Node4	Sigmoid Node5
Threshold	0.23	4.55	3.31
Sex	-1.72	-2.51	-1.45
C2	-5.58	0.52	0.16
Level6	1.93	-5.65	2.33
Weight=Medium	1.50	6.87	5.11
Weight=low	-2.13	0.63	-18.24
Weight=High	0.40	-11.99	9.78
Allergy	0.36	-4.78	-1.53
Valvular_heart	-1.80	-3.38	-5.45
Normal	4.37	-0.91	-2.39
congenital	-0.86	-4.43	-0.80

ตารางที่ 4.3 แสดงตารางข้อมูลน้ำหนักระหว่าง Hidden node และ Output node

List of output node	Sigmoid Node0 (Normal)	Sigmoid Node1 (Fat)	Sigmoid Node2 (Thin)
Threshold	-10.83	7.73	-4.02
Sigmoid Node3	4.02	-4.81	-0.14
Sigmoid Node4	6.43	-12.03	4.82
Sigmoid Node5	8.51	-4.64	-3.54

4.1 ผลลัพธ์การพยากรณ์โรคอ้วนจากข้อมูลนำเข้าภายใต้การใช้เทคนิคฟัซซีดาต้าไมนิง

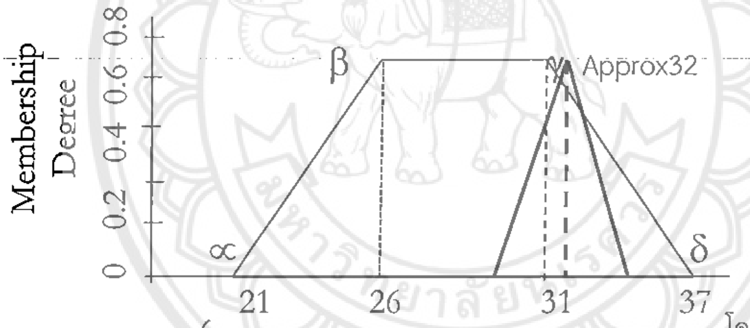
ผลลัพธ์ของการพยากรณ์โรคอ้วนจากการนำเข้าข้อมูล (ดังตารางที่ 4.4)ภายใต้การทำงานของงเทคนิคฟัซซีดาต้าไมนิง เริ่มต้นจากระบบดำเนินการสำรวจลักษณะฟัซซีแอททริบิวต์ที่จะต้องทำการปรับเปลี่ยนค่า นั่นคือ ฟัซซีแอททริบิวต์ Weight ที่มีค่าเท่ากับ “Approx32” ,ฟัซซีแอททริบิวต์ Height มีค่าเท่ากับ“Medium” และฟัซซีแอททริบิวต์ Income_parent มีค่าเท่ากับ “Medium” ตามลำดับ ทั้งนี้ภายใต้แอททริบิวต์ fat นั้นได้กำหนดให้ Thin =0, Normal=1 และ Fat =2 ตามลำดับ เพื่อประกอบในการดำเนินการที่มีรายละเอียดการทำงานของโครงข่ายประสาทเทียมที่มีลักษณะ 31-3-3 โดยใช้ข้อมูลตัวอย่างนำเข้าที่ต้องการพยากรณ์ที่มีลักษณะดังต่อไปนี้

ตารางที่ 4.4 แสดงตัวอย่างการนำเสนอข้อมูลสำหรับพยากรณ์โรคอ้วนด้วยเทคนิคฟัซซีดาต้าไมนิง

sex	c2	level6	weight	height	allergy	Valvular _heart	normal
1	0	0	Approx32	Medium	0	0	1
congenital	exercise	game	money	fast_food	soft_drink	candy	fried
0	1	0	1	2	2	2	2
aunt	Elementary _education	no_study	Contractors	Enterprise _employee	Income _parent	Fat	
0	1	0	1	0	Medium	?	

ขั้นตอนแรกจะต้องแปลงข้อมูลจากตารางนำเข้าข้อมูล โดยที่มีโหนดนำเข้าข้อมูล(Input node) จำนวน 31 โหนด มีโหนดซ่อน(Hidden node) จำนวน 3 โหนด) และมีโหนดแสดงผล (Output node) จำนวน 3 โหนด โดยงานวิจัยนี้จะแสดงรายละเอียดเพียงบางส่วนของแอททริบิวต์ที่เปรียบเทียบกับโหนดของตัวแบบจำลองที่มีลักษณะฟัซซี เช่น weight ="Approx32" เปรียบเทียบค่าระดับความใกล้เคียงกับโหนดของ “High” , “Medium”, และ “Low” ตามลำดับ ซึ่งในงานวิจัยนี้จะแสดงการเปรียบเทียบค่า Approx32 กับค่าWeight = “Medium” มีรายละเอียดดังนี้ คือ

แอททริบิวต์ weight ให้เป็น a1 จะได้ คือ



โดยที่ x คือค่าประมาณ Approx 32

α คือค่า marginของกราฟฟังก์ชันสมาชิกรูปสามเหลี่ยม

$$\mu(x, a: \alpha, \beta, \gamma, \delta) = \begin{cases} 1 \ \ when\ \beta \le x \le \gamma \\ 0 \ when\ x + a \le \alpha \ or\ x - a \ge \delta \\ \frac{\delta - x + a}{\delta - \gamma + a} \ \ when\ x > \gamma \ and\ x - a < \delta \\ \frac{x + a - \alpha}{a + \beta - \alpha} \ \ when\ x < \beta \ and\ x + a > \alpha \end{cases}$$

$$\max(h1, h2); h1 = \frac{x+a-\alpha}{(\beta-\alpha)+a}, h2 = \frac{\alpha-x-a}{(\beta-\alpha)-a} \ \ when\ x - a \ge \alpha \ and\ \alpha - x < \delta)$$

$$or\ h1 = \frac{\delta-x+a}{(\delta-\gamma)+a}, h2 = \frac{\delta-x-a}{(\delta-\gamma)-a} \ \ or\ x + a \le \delta \ and\ \gamma < x < \delta)$$

ภาพที่ 4.1 แสดงการเปรียบเทียบระหว่าง Approx 32 กับ ‘Medium’

จากภาพที่ 4.1 เราสามารถคำนวณค่าเพื่อเปรียบเทียบระหว่าง Approx32 กับ Medium โดยกำหนดให้

- - - Approx32

—— Medium

ดังนั้น $degree = \frac{\delta - x + a}{\delta - y + a}$

$= \frac{37-32+3}{37-31+3}$

= 0.89

ดังนั้น เมื่อทำการคำนวณค่าเปรียบเทียบจะพบว่า ค่าความใกล้เคียงระหว่าง Approx32 กับ "High" ,"Medium" และ " Low" จะมีค่าเท่ากับ 0.18 ,0.89 และ 0.0 ตามลำดับ ซึ่งแสดงว่าค่า Approx32จะใช้ค่าน้ำหนักของโหนดฟัซซีแอททริบิวต์ Weight = "Medium" ซึ่งมีค่าเท่ากับ Sigmoid node3 = 1.50, Sigmoid node 4 = 6.87 และ Sigmoid node5 = 5.11 ตามลำดับ เพื่อคำนวณสมการโครงข่ายประสาทเทียมในลำดับถัดไป ซึ่งขั้นตอนต่อไปเป็นการแปลงข้อมูลนำเข้าสำหรับทดสอบแบบจำลองโครงข่ายประสาทเทียม ตัวอย่าง ฟัซซีแอททริบิวต์ weight = "Approx32" มีรายละเอียดดังนี้ คือ

	Sex	C2	Level6	weight	height	allergy	Valvular_heart	Normal
Range	0.50	0.50	0.50	31.50	33	0.50	0.50	0.50
Base	0.50	0.50	0.50	45.50	139	0.50	0.50	0.50
Norm	1.00	-1.00	-1.00	-0.43	-0.21	-1.00	-1.00	1.00

	congenital	exercise	game	money	Fast_food	Soft_drink	candy	fried
Range	0.50	1.50	0.50	1.00	1.50	1.50	1.50	1.50
Base	0.50	1.50	0.50	2.00	1.50	1.50	1.50	1.50
Norm	-1.00	-0.33	-1.00	-1.00	0.33	0.33	0.33	0.33

	aunt	Elementary_education	No_study	contractors	Enterprise_employee	Income_parent
Range	0.50	0.50	0.50	0.50	0.50	1.00
Base	0.50	0.50	0.50	0.50	0.50	2.00
norm	-1.00	1.00	-1.00	1.00	-1.00	0.00

จะคำนวณค่าแอททริบิวต์ต่างๆ ที่แปลงแล้วมาใส่ในสมการเพื่อส่งไปยัง active function ซึ่งเป็นโหนดใน Hidden layer มีรายละเอียดดังนี้ คือ

Sigmoid node3 = x_1 = 1.00x(-1.72)-1.00x(-5.58)-1.00x(1.93)-0.43x(1.50)-0.21x(6.93)-
1.00x(0.36)-1.00x(-1.80)-1.00x(4.37)-1.00x(-0.86)-0.33x(6.71)-1.00x(-5.16)-
1.00x(4.00)+0.33x(0.81)+0.33x(-10.79)+0.33x(-6.85)+0.33x(-12.25)
-1.00x(-3.46)+1.00x(-4.88)-1.00x(2.35)+1.00x(-5.53)-1.00x(-2.63)
+0.00(9.77)+0.23
= -19.33

$f(x_1) = \frac{1}{1+e^{-x_1}} = \frac{1}{1+e^{19.33}} = 0$

Sigmoid node4= x_2=1.00x(-2.51)-1.00x(0.52)-1.00x(-5.65)-0.43x(6.87)-0.21x(0.24)-1.00x(-4.78)
-1.00x(-3.38)-1.00x(-0.91)-1.00x(-4.43)-0.33x(-4.73)-1.00x(4.13)
-1.00x(-1.23)+0.33x(-1.41)+0.33x(-5.43)+0.33x(-1.46)+0.33x(-7.98)
-1.00x(-5.50)+1.00x(1.93)-1.00x(5.04)+1.00x(1.08)-1.00x(3.01)+0.00(1.92)
+4.55
= 11.42

$f(x_2) = \frac{1}{1+e^{-x_2}} = \frac{1}{1+e^{-11.42}} = 1$

Sigmoid node5= x_3=1.00x(-1.45)-1.00x(0.16)-1.00x(2.33)-0.43x(5.11)-0.21x(-1.01) -1.00x(-1.53)
-1.00x(-5.45)-1.00x(-2.39)-1.00x(-0.80)-0.33x(7.69)-1.00x(5.93)-1.00x(-1.29)
+0.33x(3.86)+0.33x(-10.30)+0.33x(6.35)+0.33x(-8.07)-1.00x(-2.02)
+1.00x(-5.70)-1.00x(-5.26)+1.00x(5.39)-1.00x(0.81)+0.00(-0.75)+3.31
= -0.19

$f(x_3) = \frac{1}{1+e^{-x_3}} = \frac{1}{1+e^{0.19}} = 0.45$

หลังจากนั้นจะเข้าสู่โหนดการคำนวณเพื่อนำเสนอผลลัพธ์ ภายใต้การแสดงผลของโหนด 3 โหนด ซึ่งมีรายละเอียดนั่นคือ

Sigmoid node0= x_0 (โหนด Fat=Normal)= (0*4.02)+(1*6.43)+(0.45*8.51)-10.83 = -0.57

$f(x_0) = \frac{1}{1+e^{-x_0}} = \frac{1}{1+e^{0.57}} = 0.36$

Sigmoid node1 = x_1 (โหนด Fat=Fat) = (0*-14.81)+(1*-12.03)+(0.45*-4.64)+7.73 = -6.39

$f(x_1) = \frac{1}{1+e^{-x_1}} = \frac{1}{1+e^{6.39}} = 0$

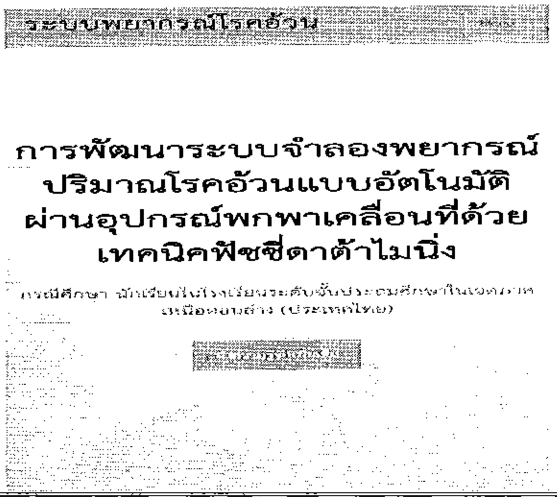
Sigmoid node2 = x_2 (โหนด Fat=Thin) = (0*-0.14) +(1*4.82)+(0.45*-3.54) -4.02 = -0.79

$f(x_2) = \frac{1}{1+e^{-x_2}} = \frac{1}{1+e^{0.79}} = 0.31$

ดังนั้น จากค่าที่คำนวณสามารถสรุปได้ว่า ข้อมูลตัวอย่างที่นำเข้าสู่ระบบพยากรณ์ ระบบสามารถพยากรณ์ได้ว่าค่าของแอททริบิวต์ Fat เข้าใกล้Sigmoid node0= x_0 ซึ่ง กำหนดให้เป็นระดับ “**Normal**” เนื่องจากผลลัพธ์ที่ได้โดยระบบสามารถดำเนินการซึ่งมีค่าสูงสุดนั่นคือ 0.36

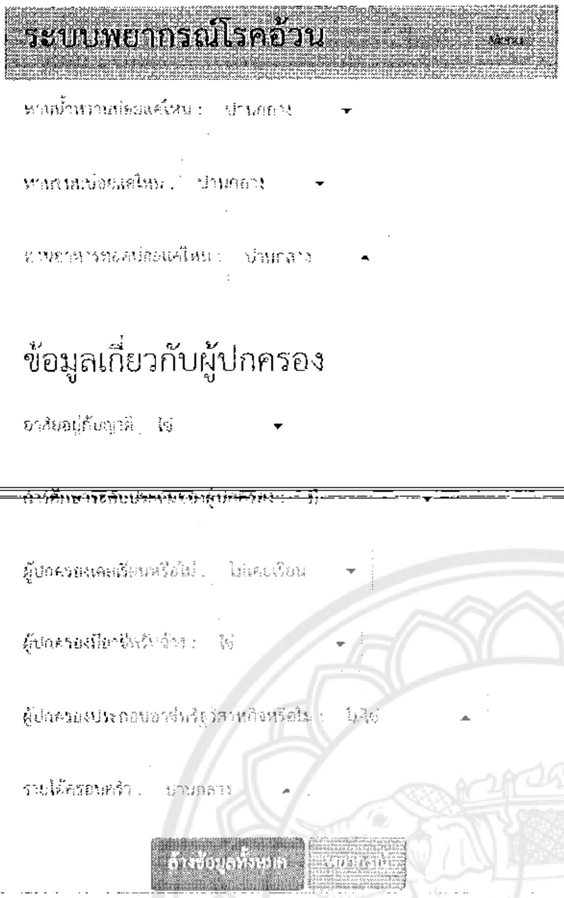
4.2 ตัวอย่างการทำงานของแอพพลิเคชันที่ได้พัฒนา

การทำงานของระบบโปรแกรมจะกำหนดให้ผู้ใช้กรอกข้อมูลรายละเอียดของเด็กนักเรียนแต่บะคนได้เพื่อทำนายว่านักเรียนคนนี้เมื่อเข้าเรียนที่โรงเรียนต่างๆ โดยอาศัยปัจจัยของครอบครัวนักเรียนแสภาพโรงเรียน เพื่อใช้พยากรณ์การเกิดโรคอ้วนของนักเรียน ซึ่งการทำงานมีรายละเอียดเริ่มต้น (ดังภาพที่ 4.2) และเมื่อทดลองกรอกข้อมูล (ดังภาพที่ 4.3) จะพบว่าผลการพยากรณ์นั้นจะแสดงผลว่า นักเรียนที่มีรายละเอียดดังกล่าวที่ได้กรอกเข้าไปนั้น จะมีลักษณะ normal นั่นคือ ไม่เป็นโรคอ้วน (สถานะปกติ) (ดังภาพที่ 4.4)

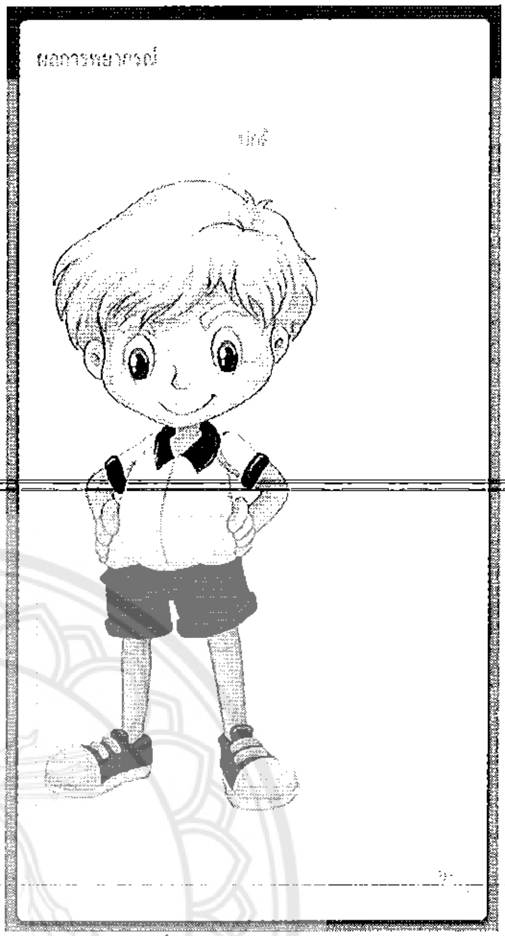


ภาพที่ 4.2 แสดงรายละเอียดเริ่มต้นเข้าสู่การทำงานของโปรแกรม

ภาพที่ 4.3(1)แสดงรายละเอียดบางส่วนของการกรอกข้อมูล



ภาพที่ 4.3(2)แสดงรายละเอียดบางส่วนการกรอกข้อมูล



ภาพที่ 4.4 แสดงผลการพยากรณ์

สรุปผลการวิเคราะห์และวิจารณ์

ปกติการวัดความอ้วนจะดำเนินการได้จากการคำนวณค่าดัชนีมวลกาย แต่ในงานวิจัยนี้พบว่า การวัดค่าด้วยหลักการดังกล่าวนั้นไม่สามารถจะทำได้ หากข้อมูลนำเข้าของ น้ำหนัก และส่วนสูงนั้นเนข้อมูลในลักษณะเชิงฟัซซี ผู้วิจัยจึงได้ดำเนินการนำเสนอวิธีการด้วยการดำเนินวิจัยภายใต้หัวข้อเรื่อง การพัฒนาระบบจำลองพยากรณ์ปริมาณโรคอ้วนแบบอัตโนมัติผ่านอุปกรณ์พกพาเคลื่อนที่ด้วยเทคนิคฟัซซี่ดาต้าไมนิ่ง: กรณีศึกษานักเรียนในโรงเรียนระดับชั้นประถมศึกษาในเขตจังหวัดภาคเหนือตอนล่าง (ประเทศไทย) ทั้งนี้จากการดำเนินวิจัยเราจะพบว่างานวิจัยนี้สามารถนำเสนอข้อมูลนำเข้าจากปัจจัยเชิงฟัซซีได้ โดยแนวทางของการพัฒนางานนี้จะให้ความสำคัญของการนำเข้าข้อมูล ซึ่งในงานวิจัยได้กำหนดปัจจัยหลักๆ ที่มีลักษณะเชิงฟัซซี ได้แก่ น้ำหนัก ส่วนสูงและ รายได้ของครอบครัว ตามลำดับ เพื่อใช้สำหรับเป็นส่วนหนึ่งของการสร้างตัวแบบจำลองพยากรณ์โรคอ้วนที่มีการอาศัยปัจจัยอื่นๆ ร่วมด้วย โดยในงานวิจัยยังได้มีการนำเสนอรูปแบบของการตรวจสอบปัจจัยที่มีผลต่อการเกิดโรคอ้วนด้วยหลักการเชิงสถิติศาสตร์ภายใต้การใช้เทคนิคเพียร์สัน พร้อมทั้งมีการกำหนดลักษณะข้อมูลเชิงคุณภาพที่มีอยู่ให้อยู่ในรูปแบบของการทำดัมมี หลักจากนั้นจะพบว่าตัวแบบพยากรณ์ที่นำมาใช้สำหรับสร้างแอพพลิเคชันนั้นได้จากการเปรียบเทียบการทำงานของอัลกอริธึมโครงข่ายประสาทเทียมและต้นไม้ช่วยตัดสินใจ ซึ่งจะพบว่า การทำงานของภูครงข่ายประสาทเทียมรูปแบบ 31-3-3 ที่มีการจัดการแบ่งกลุ่มข้อมูลในลักษณะ cross validation fold-10 มีความเหมาะสมมากที่สุด อย่างไรก็ตามการงานวิจัยนี้ปัญหาหลักของการดำเนินการนั้น การรวบรวมข้อมูลเพื่อให้เกิดความหลากหลายจากระดับชั้นปีของนักเรียนระดับชั้นประถมศึกษานั้น จะพบว่าส่วนใหญ่จะได้รับความร่วมมือจากโรงเรียนระดับเขตนอกเมืองเป็นหลักและเป็นส่วนของภาครัฐบาล ส่วนข้อมูลของนักเรียนในเขตอำเภอเมืองทั้งส่วนของภาครัฐและภาคเอกชนนั้น จะได้รับความร่วมมือค่อนข้างน้อยถึงน้อยมาก จึงทำให้ข้อมูลที่ได้สำหรับการสร้างตัวแบบพยากรณ์อาจจะมีความคลาดเคลื่อนต่อการนำไปใช้งานสำหรับพยากรณ์การเกิดโรคอ้วนในระดับหนึ่ง ซึ่งแอพพลิเคชั่นที่พัฒนาจากงานวิจัยนี้ เขตที่ตั้งของโรงเรียนหรือแม้แต่ขนาดของโรงเรียนก็เป็นส่วนหนึ่งของปัจจัยที่สำคัญต่อการเกิดโรคอ้วนในเด้กนักเรียนระดับชั้นประถมศึกษาที่มีอายุระหว่าง 6-12 ปีตามเกณฑ์มาตรฐาน

บรรณานุกรม

[1] ชนวัฒน์ ศรีสอ้าน, ฐานข้อมูล คลังข้อมูลและเหมืองข้อมูล

[2] ทักษพล ธรรมรังสี, สงขลานครินทร์เวชสาร ปีที่29 ฉบับที่2 มีนาคม-เมษายน 2554

[3] ธนาวุฒิ ประกอบผล,โครงข่ายประสาทเทียม,วารสาร มฉก. วิชาการ ปีที่12 ฉบับที่24 มกราคม-มิถุนายน 2552

[4] วิชัย เทียนถาวร ,หนังสือพิมพ์มติชน ฉบับวันที่1 เมษายน 2556

[5] วิวัฒน์ ชินนาทศิริกุล, การทำเหมืองข้อมูลด้วยการใช้โปรแกรม weka หาความสัมพันธ์ของข้อมูล ,วารสารการเวก ฉบับวันเจ้าฟ้าปีที่7 ฉบับที่1 เดือนกุมภาพันธ์ 2554

[6] หนังสือพิมพ์ฐานเศรษฐกิจ ฉบับวันที่ 9 - 11 พ.ค. 2556

[7] Chittayasothorn S., Toward Fuzzy Temporal Databases with Temporal Fuzzy Linguistic Terms, The second International Conference on the Applications of Digital Information and Web Technologies (ICADIWT2009) August 4-6, 2009 IEEE.

[8] Holland. 1975. Adaptation in Natural and Artificial System. The University of Michigan Press. 54(2): 186-198.

[9] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Introduction to Data Mining, Addison-Wesley, 2006.

[10] Richard J. Roiger, Michael W. Geatz ,Data Mining A tutorial-Based Primer, Pearson Education Inc., 2003.

[11] Riordan D., Hansen B. K., A Fuzzy Case-Based System forWeather Prediction, Engineering intelligent systems, September,2002.

[12] โรคอ้วน ,สืบค้นวันที่ 15 พฤษภาคม 2557 แหล่งข้อมูล: http://www.gpo.or.th /rdi/html/ obes.html

13] สืบค้นวันที่ 20 พฤษภาคม 2557 แหล่งข้อมูล : http: http//www.banmuang.co.th/2013/02เด็ก ประถมเสี่ยงโรคอ้วน ประจำวันวันอาทิตย์ที่ 18 พฤษภาคม พ.ศ. 2557

[14] สืบค้นเมื่อวันที่ 23 พฤษภาคม 2557 แหล่งข้อมูล: http://www4.thaihealth.or.th/healthcontent /special_report/23934 ประจำวันที่ 18 สิงหาคม 2554

ภาคผนวก ก
การคำนวณเปรียบเทียบการใช้แอพพลิเคชั่น

ในส่วนของภาคผนวค ก ได้ดำเนินการทดสอบตัวอย่างข้อมูลจริงจำนวน 5 ตัวอย่างเพื่อทดสอบการคำนวณด้วยแอพพลิเคชั่นของระบบว่า มีค่าความผิดพลาดของการคำนวณในการพยากรณ์โรคอ้วนในระดับใด ดดยมีรายละเอียดของข้อมูลนำเข้ามนลักษณะต่างๆ ดังตารางผนวก ก-1

ตารางที่ผนวก ก-1 แสดงตัวอย่างการนำเสนอข้อมูลสำหรับพยาการณ์โรคอ้วนด้วยเทคนิคฟัซซีดาต้าไมนิง

แอททริบิวต์	รายละเอียดนำเข้าข้อมูลของนักเรียน				
	คนที่ 1	คนที่ 2	คนที่ 3	คนที่ 4	คนที่ 5
Sex	1	1	1	1	1
C2	0	0	0	0	0
Level6	0	0	0	0	0
Weight	Medium	Low	30	55	Low
Height	Low	120	Medium	Low	125
Allergy	0	0	0	0	0
Valvular_heart	0	0	0	0	0
Normal	1	1	1	1	1
Congenital	0	0	0	0	0
Exercise	1	1	1	2	1
Game	0	0	0	0	0
Money	1	1	1	1	1
Fast_food	1	1	2	2	1
Soft_drink	1	1	2	2	1
Candy	1	1	2	2	1
Fried	1	1	2	2	1
Aunt	0	0	0	0	0
Elementary_education	1	1	1	1	1
No_study	0	0	0	0	0
Contractors	1	0	1	1	0
Enterprise_employee	0	0	0	0	0
Income_parent	Medium	High	Medium	Medium	High
Fat	Normal (1)	Normal (1)	Normal(1)	Fat(2)	Thin(0)
Fat(application prediction)	Normal	Normal	Thin	Fat	Thin

ขั้นตอนแรกจะต้องแปลงข้อมูลจากตารางนำเข้าข้อมูล ของนักเรียนคนที่ 1 โดยที่มีโหนดนำเข้าข้อมูล(Input node) จำนวน 28 โหนด มีโหนดซ่อน(Hidden node) จำนวน 3 โหนด และมีโหนดแสดงผล (Output node) จำนวน 3 โหนด โดยงานวิจัยนี้จะแสดงรายละเอียดเพียงบางส่วนของแอททริบิวต์ที่เปรียบเทียบกับโหนดของตัวแบบจำลองที่มีลักษณะฟัซซี เช่น weight คือ "Medium" สำหรับค่า Height คือ"Low", และ

ค่าของ Income_parent คือ “Medium” ตามลำดับ ดังนั้นทำการเปรียบเทียบค่าระหว่างฟัซซีแอททริบิวต์ดังกล่าว ซึ่งมีรายละเอียดคือ

ค่า Weight มีค่า “Medium”ปรียบเทียบกับค่า”Medium” นั่นคือ = 1

Weight มีค่า “Medium” เปรียบเทียบกับค่า “High” มีรายละเอียดการคำนวณดังต่อไปนี้

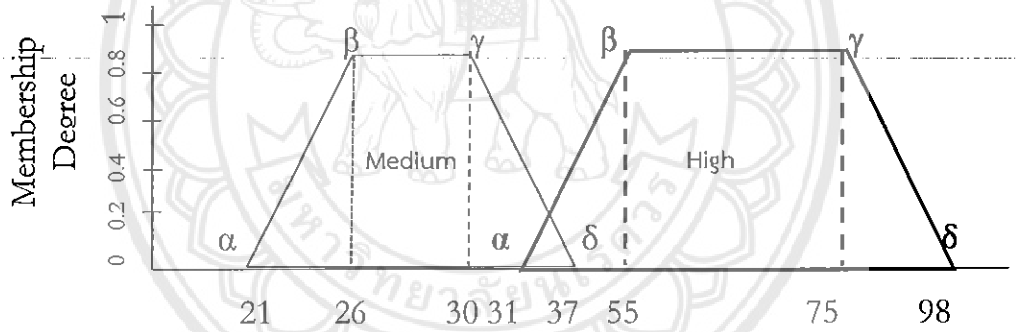
รูปแบบการทำ Fuzzy attribute matchingในลักษณะค่า พจน์ภาษาฟัซซี (Fuzzy linguistic term) กับ พจน์ภาษาฟัซซี (Fuzzy linguistic term) จะทำการเปรียบเทียบกันระหว่างค่าพจน์ภาษาฟัซซี *2* ค่าใดๆ ในแอทริบิวต์เดียวกัน ซึ่งวิธีการหาค่าดีกรีความคล้ายคลึง ระหว่าง *2* ฟัซซีเซ็ตใดๆ โดยเราจะได้ผลลัพธ์ความคล้ายคลึงของค่าดีกรีระหว่างคู่พจน์ภาษาฟัซซีที่มีค่ามากกว่าหรือเท่ากับค่า t (Threshold)

$$S = k/(\frac{w_1-w_2}{\delta_1-\alpha_2} - k) \quad \text{โดยที่ } k - \frac{\delta_1-\alpha_2}{\delta_1-\gamma_1+\delta_2-\gamma_2}$$

$$w_i = w_{bi} + w_{ti}\,;\, i = 1,2$$

ความกว้างด้านล่าง (bottom width) $w_{bi} = \frac{(\delta_i-\alpha_i)}{2}$

ความกว้างด้านบน (top width) $w_{ti} = \frac{(\gamma_i-\beta_i)}{2}$



ภาพ ผนวค ก-1 แสดงระดับความสัมพันธ์ระหว่างฟัซซีแอททริบิวต์ weight ค่าระหว่าง Medium กับ High

ดังนั้น เมื่อทำการคำนวณค่าเปรียบเทียบจะพบว่า ค่าความใกล้เคียงระหว่าง “High” กับ”Medium” จะมีค่าเท่ากับ 0.05 ซึ่งถือได้ว่าคล้ายคลึงกันน้อยมาก

ขั้นตอนการทดสอบแบบจำลองสามารถกระทำได้โดยทดสอบจากข้อมูลนำเข้าจำนวน 5 ตัวอย่างมีรายละเอียดดังนี้

ตัวอย่างของข้อมูล**<u>นักเรียนคนที่ 1</u>**

- ขั้นตอนแรกจะต้องแปลงข้อมูลจากตารางนำเข้า โดยมีรายละเอียดดังนี้ คือ

  range=(max-min)/2 โดย max และ min ได้จากค่ามากสุดและน้อยสุดของแต่ละแอททริบิวต์

  base=(max+min)/2

  norm_attribute= (attribute-base)/range

ดังนั้น แอททริบิวต์ Sex ให้ปีน a1 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a1 =(1-0.50)/0.50 = 1.00

ดังนั้น แอททริบิวต์ c2 ให้เป็น a2 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a2 =(1-0.50)/0.50 = 1.00

ดังนั้น แอททริบิวต์ Level6 ให้เป็น a3 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a3 =(1-0.50)/0.50 = 1.00

ดังนั้น แอททริบิวต์ weight ให้เป็น a4 จะได้ คือ

range =(2-0)/2 =1.00

base =(2+0)/2 = 1.00

norm_a4 =(1-1)/1 = 0.00

ดังนั้น แอททริบิวต์ height ให้เป็น a5 จะได้ คือ

range =(2-0)/2 =1.00

base =(2+0)/2 =1.00

norm_a5 =(0-1)/1 = -1.00

ดังนั้น แอททริบิวต์ allergy ให้เป็น a6 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a6 =(0-0.50)/0.50 = -1.00

ดังนั้น แอททริบิวต์ valvular_heart ให้เป็น a7 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a7 =(0-0.50)/0.50 = -1.00

ดังนั้น แอททริบิวต์ normal ให้เป็น a8 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a8 =(1-0.50)/0.50 = 1.00

ดังนั้น แอททริบิวต์ congenital ให้เป็น a9 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a9 =(0-0.50)/0.50 = -1.00

ดังนั้น แอททริบิวต์ exerciseให้เป็น a10 จะได้ คือ

range =(3-0)/2 =1.50

base =(3+0)/2 =1.50

norm_a10 =(1-1.50)/1.50=-0.33

ดังนั้น แอททริบิวต์ game ให้เป็น a11 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a11 =(0-0.50)/0.50 = -1.00

ดังนั้น แอททริบิวต์ money ให้เป็น a12 จะได้ คือ

range =(3-1)/2 =1.00

base =(3+1)/2=2.00

norm_a12 =(1-2)/1 = -1.00

ดังนั้น แอททริบิวต์ fast_food ให้เป็น a13 จะได้ คือ

range =(3-0)/2 =1.50

base =(3+0)/2=1.50

norm_a13 =(1-1.50)/1.50 =-0.33

ดังนั้น แอททริบิวต์ soft_drink ให้เป็น a14 จะได้ คือ

range =(3-0)/2 =1.50

base =(3+0)/2=1.50

norm_a13 =(1-1.50)/1.50 =-0.33

ดังนั้น แอททริบิวต์ candy ให้เป็น a15 จะได้ คือ

range =(3-0)/2 =1.50

base =(3+0)/2=1.50

norm_a13 =(1-1.50)/1.50 =-0.33

ดังนั้น แอททริบิวต์ fried ให้เป็น a16 จะได้ คือ

range =(3-0)/2 =1.50

base =(3+0)/2=1.50

norm_a13 =(1-1.50)/1.50 =-0.33

ดังนั้น แอททริบิวต์ aunt ให้เป็น a17 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a17 =(0-0.50)/0.50 = -1.00

ดังนั้น แอททริบิวต์ elementary_education ให้เป็น a18 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a18 =(1-0.50)/0.50=1.00

ดังนั้น แอททริบิวต์ no_study ให้เป็น a19 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a19 =(0-0.50)/0.50 = -1.00

ดังนั้น แอททริบิวต์ contractors ให้เป็น a20 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a20 =(1-0.50)/0.50=1.00

ดังนั้น แอททริบิวต์ enterprise_employee ให้เป็น a21 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a21 =(0-0.50)/0.50 = -1.00

ดังนั้น แอททริบิวต์ income_parent ให้เป็น a22 จะได้ คือ

range =(2-0)/2 =1.00

base =(2+0)/2=1.00

norm_a22 =(1-1)/1 = 0.00

	Sex					C2				
	นร.1	นร.2	นร.3	นร.4	นร.5	นร.1	นร.2	นร.3	นร.4	นร.5
Range	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Base	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Norm	1.00	1.00	1.00	1.00	1.00	-1.00	-1.00	-1.00	-1.00	-1.00

	Level6					Weight				
	นร.1	นร.2	นร.3	นร.4	นร.5	นร.1	นร.2	นร.3	นร.4	นร.5
Range	0.50	0.50	0.50	0.50	0.50	1.00	1.00	1.00	1.00	1.00
Base	0.50	0.50	0.50	0.50	0.50	1.00	1.00	1.00	1.00	1.00
Norm	-1.00	-1.00	-1.00	-1.00	-1.00	0.00	-1.00	0.00	1.00	-1.00

	height					Allergy				
	นร.1	นร.2	นร.3	นร.4	นร.5	นร.1	นร.2	นร.3	นร.4	นร.5
Range	1.00	1.00	1.00	1.00	1.00	0.50	0.50	0.50	0.50	0.50
Base	1.00	1.00	1.00	1.00	1.00	0.50	0.50	0.50	0.50	0.50
Norm	-1.00	-0.50	0.00	-1.00	0.00	-1.00	-1.00	-1.00	-1.00	-1.00

	Valvular_heart					Normal				
	นร.1	นร.2	นร.3	นร.4	นร.5	นร.1	นร.2	นร.3	นร.4	นร.5
Range	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Base	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Norm	-1.00	-1.00	-1.00	-1.00	-1.00	1.00	1.00	1.00	1.00	1.00

	congenital					Exercise				
	นร.1	นร.2	นร.3	นร.4	นร.5	นร.1	นร.2	นร.3	นร.4	นร.5
Range	0.50	0.50	0.50	0.50	0.50	1.50	1.50	1.50	1.5	1.50
Base	0.50	0.50	0.50	0.50	0.50	1.50	1.50	1.50	1.5	1.50
Norm	-1.00	-1.00	-1.00	-1.00	-1.00	-0.33	-0.33	-0.33	0.33	-0.33

	game					Money				
	นร.1	นร.2	นร.3	นร.4	นร.5	นร.1	นร.2	นร.3	นร.4	นร.5
Range	0.50	0.50	0.50	0.50	0.50	1.00	1.00	1.00	1.00	1.00
Base	0.50	0.50	0.50	0.50	0.50	2.00	2.00	2.00	2.00	2.00
Norm	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00

	Fast_food					Soft_drink				
	นร.1	นร.2	นร.3	นร.4	นร.5	นร.1	นร.2	นร.3	นร.4	นร.5
Range	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Base	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Norm	-0.33	-0.33	0.33	0.33	-0.33	-0.33	-0.33	0.33	0.33	-0.33

	candy					Fried				
	นร.1	นร.2	นร.3	นร.4	นร.5	นร.1	นร.2	นร.3	นร.4	นร.5
Range	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Base	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Norm	-0.33	-0.33	0.33	0.33	-0.33	-0.33	-0.33	0.33	0.33	-0.33

	aunt					Elementary_education				
	นร.1	นร.2	นร.3	นร.4	นร.5	นร.1	นร.2	นร.3	นร.4	นร.5
Range	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Base	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
norm	-1.00	-1.00	-1.00	-1.00	-1.00	1.00	1.00	1.00	1.00	1.00

	No_study					Contractors				
	นร.1	นร.2	นร.3	นร.4	นร.5	นร.1	นร.2	นร.3	นร.4	นร.5
Range	0.50	0.50	0.50	0.50	0.50	0.50	0.5	0.50	0.50	0.5
Base	0.50	0.50	0.50	0.50	0.50	0.50	0.5	0.50	0.50	0.5
norm	-1.00	-1.00	-1.00	-1.00	-1.00	1.00	-1.00	1.00	1.00	-1.00

	Enterprise_employee					Income_parent				
	นร.1	นร.2	นร.3	นร.4	นร.5	นร.1	นร.2	นร.3	นร.4	นร.5
Range	0.50	0.50	0.50	0.50	0.50	1.00	1.00	1.00	1.00	1.00
Base	0.50	0.50	0.50	0.50	0.50	2.00	1.00	2.00	2.00	1.00
norm	-1.00	-1.00	-1.00	-1.00	-1.00	0.00	0.00	0.00	0.00	1.00

จะคำนวณค่าแอททริบิวต์ต่างๆ ที่แปลงแล้วมาใส่ในสมการเพื่อส่งไปยัง active function ซึ่งเป็นโหนดใน Hidden layer มีรายละเอียดดังนี้ คือ

Sigmoid node3 = x_1 = 1.00x(-1.72)-1.00x(-5.58)-1.00x(1.93)+0x(1.50)-1.00x(6.93)-1.00x(0.36)-1.00x(-1.80)+1.00x(4.37)-1.00x(-0.86)-0.33x(6.71)-1.00x(-5.16)-1.00x(4.00)-0.33x(0.81)-0.33x(-10.79)-0.33x(-6.85)-0.33x(-12.25)-1.00x(-3.46)+1.00x(-4.88)-1.00x(2.35)+1.00x(-5.53)-1.00x(-2.63)+0x(9.77)+0.23

= 2.05

$$f(x_1) = \frac{1}{1+e^{-x_1}} = \frac{1}{1+e^{-2.05}} = 0.89$$

Sigmoid node4= x_2=1.00x(-2.51)-1.00x(0.52)-1.00x(-5.65)+0x(6.87)-1.00x(0.24)-1.00x(-4.78)
-1.00x(-3.38)+1.00x(-0.91)-1.00x(-4.43)-0.33x(-4.73)-1.00x(4.13)
-1.00x(-1.23)-0.33x(-1.41)-0.33x(-5.43)-0.33x(-1.46)-0.33x(-7.98)
-1.00x(-5.50)+1.00x(1.93)-1.00x(5.04)+1.00x(1.08)-1.00x(3.01)+0x(1.92)
+4.55
= 23.10

$$f(x_2) = \frac{1}{1+e^{-x_2}} = \frac{1}{1+e^{-23.1}} = 1$$

Sigmoid node5= x_3=1.00x(-1.45)-1.00x(0.16)-1.00x(2.33)+0x(5.11)-0x(-1.01) -1.00x(-1.53)
-1.00x(-5.45)+1.00x(-2.39)-1.00x(-0.80)-0.33x(7.69)-1.00x(5.93)-1.00x(-1.29)
-0.33x(3.86)-0.33x(-10.30)-0.33x(6.35)-0.33x(-8.07)-1.00x(-2.02)
+1.00x(-5.70)-1.00x(-5.26)+1.00x(5.39)-1.00x(0.81)+0x(-0.75)+3.31
= 7.44

$$f(x_3) = \frac{1}{1+e^{-x_3}} = \frac{1}{1-e^{-7.44}} = 0.99$$

หลังจากนั้นจะเข้าสู่โหนดการคำนวณเพื่อนำเสนอผลลัพธ์ ภายใต้การแสดงผลของโหนด 3 โหนด ซึ่งมีรายละเอียดนั่นคือ

Sigmoid node0= x_0 (โหนด Fat=Normal)= (0.89*4.02)+(1*6.43)+(0.99*8.51)-10.83 = 7.60

$$f(x_0) = \frac{1}{1+e^{-x_0}} = \frac{1}{1+e^{-7.6}} = 0.17$$

Sigmoid node1 = x_1 (โหนด Fat=Fat) = (0.89*-14.81)+(1*-12.03)+(0.99*-4.64)+7.73 = -22.07

$$f(x_1) = \frac{1}{1+e^{-x_1}} = \frac{1}{1+e^{22.07}} = 0$$

Sigmoid node2 = x_2 (โหนด Fat=Thin) = (0.89*-0.14) +(1*4.82)+(0.99*-3.54) -4.02 = -2.83

$$f(x_2) = \frac{1}{1+e^{-x_2}} = \frac{1}{1+e^{2.83}} = 0.06$$

ดังนั้น จากค่าที่คำนวณสามารถสรุปได้ว่า ข้อมูลตัวอย่างที่นำเข้าสู่ระบบพยากรณ์ ระบบสามารถพยากรณ์ได้ว่าค่าของแอททริบิวต์ Fat เข้าใกล้Sigmoid node0= x_0 ซึ่ง กำหนดให้เป็นระดับ "Normal" เนื่องจากผลลัพธ์ที่ได้โดยระบบสามารถดำเนินการซึ่งมีค่าสูงสุดนั่นคือ 0.17

ภาคผนวก ข

ตารางแสดงรหัสข้อมูลนำเข้าในระบบแอพพลิเคชั่น

sex	c2	level6	weight	height	allergy	Valvular_heart	normal
0/1	0/1	0/1	0/1/2	0/1/2	0/1	0/1	0/1

congenital	exercise	game	money	fast_food	soft_drink	candy	fried
0/1	0/1/2/3	0/1	1/2/3	0/1/2/3	0/1/2/3	0/1/2/3	0/1/2/3

aunt	Elementary _education	no_study	Contractors	Enterprise _employee	Income _parent	Fat
0/1	0/1	0/1	0/1	0/1	1/2/3	Normal/thin/Fat

Sex	เพศ	money	เงินที่ได้มาโรงเรียน
c2 ล	เขตตำบล	ast_food	ฟาสต์ฟู้ด
level6	ระดับชั้นประถมศึกษาปีที่ 6	soft_drink	น้ำอัดลม
weight	น้ำหนัก	candy	ขนมหวาน
height	ส่วนสูง	fried	อาหารทอด
allergy	ภูมิแพ้	aunt	ญาติ (ป้า)
valvular_heart	ลิ้นหัวใจรั่ว	elementary_education	ประถมศึกษา
normal	ไม่มีโรคประจำตัว	no_study	ไม่เรียน
congenital	กรรมพันธุ์ (มีโรคอ้วนหรือไม่)	contractors รับจ้าง	รับจ้าง
exercise	ออกกำลังกาย	enterprise_employee	รัฐวิสาหกิจ
game = เล่นเกมส์	เล่นเกม	income_parent	รายได้ครอบครัว

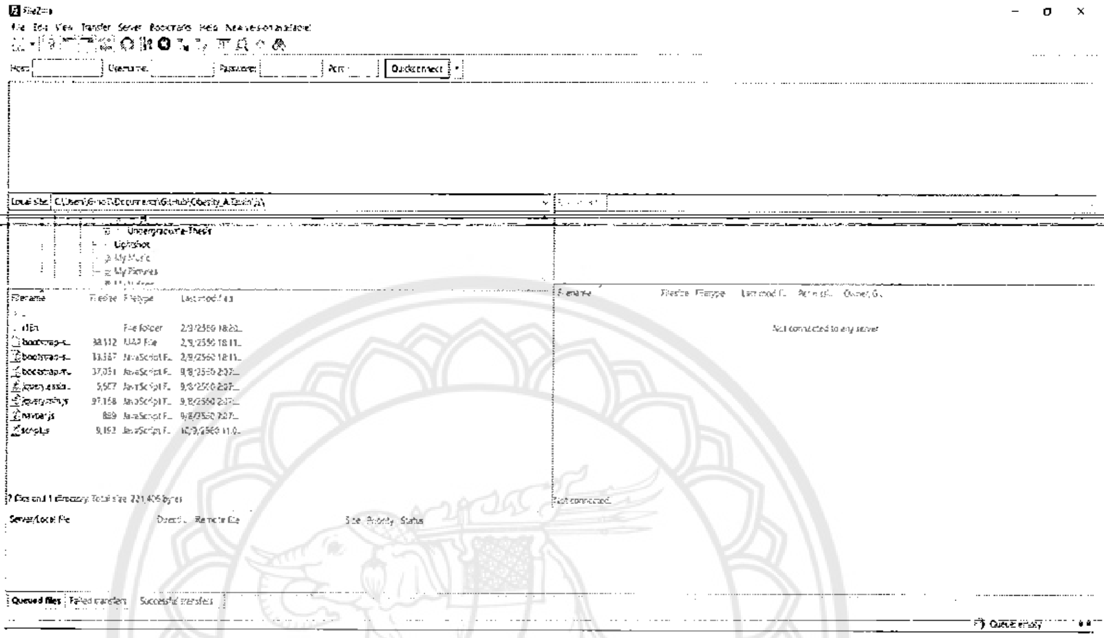
ภาคผนวก ค

ตัวอย่างโค้ดปรแกรม

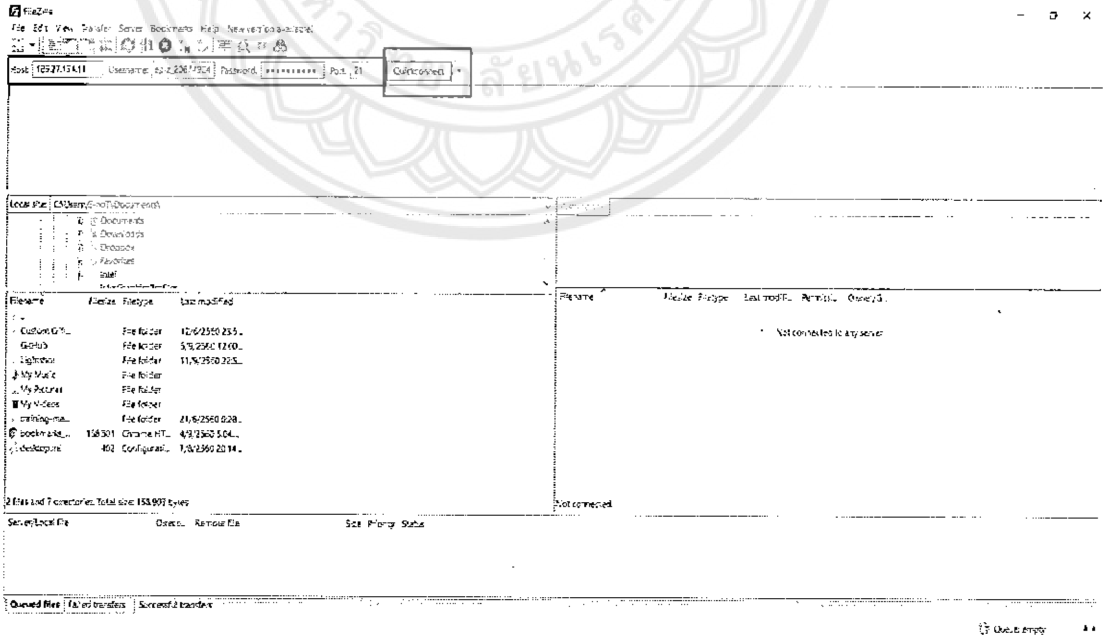
```
!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="utf-8">
    <meta http-equiv="X-UA-Compatible" content="IE=edge">
    <meta name="viewport" content="width=device-width, initial-scale=1">
    <title>ระบบพยากรณ์โรคอ้วน</title>
    <link href="./css/bootstrap.min.css" rel="stylesheet">
    <link href="./css/style.css" rel="stylesheet">
    <link href="./css/navbar.css" rel="stylesheet">
    <link href="./css/fonts.css" rel="stylesheet">
</head>

<body>
    <div id="home" data-spy="scroll" data-target=".navbar-fixed-top">
        <nav id="mainNav" class="navbar navbar-default navbar-fixed-top">
            <div class="container">
                <div class="navbar-header">
                    <button type="button" class="navbar-toggle collapsed" data-
toggle="collapse" data-target="#bs-example-navbar-collapse-1">
        <span class="sr-only">Toggle navigation</span> Menu <i class="fa fa-bars"></i>
                    </button>
            <a class="navbar-brand page-scroll" href="#home"> ระบบพยากรณ์โรคอ้วน</a>
                </div>
                <div class="collapse navbar-collapse" id="bs-example-navbar-collapse-1">
                    <ul class="nav navbar-nav navbar-right">
                        <li>
                            <a class="page-scroll" href="#home">หน้าหลัก</a>
                        </li>
                        <li>
                            <a class="page-scroll nav-font-tag" href="predict.html">ทำนาย</a>
                        .. .
                        .. .
                            <a class="page-scroll nav-font-tag" href="about.html">ผู้จัดทำ</a>
                        </li>
```

ภาคผนวก ง
การ upload file สำหรับใช้งานบนเว็บ

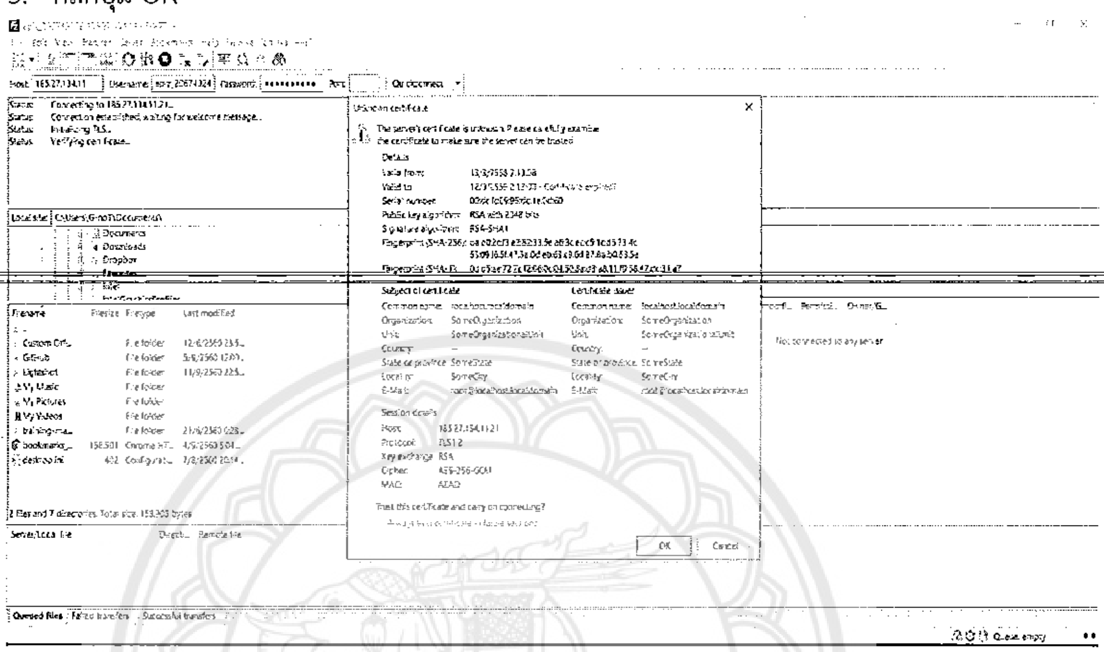
1. เปิดโปรแกรม File Zilla



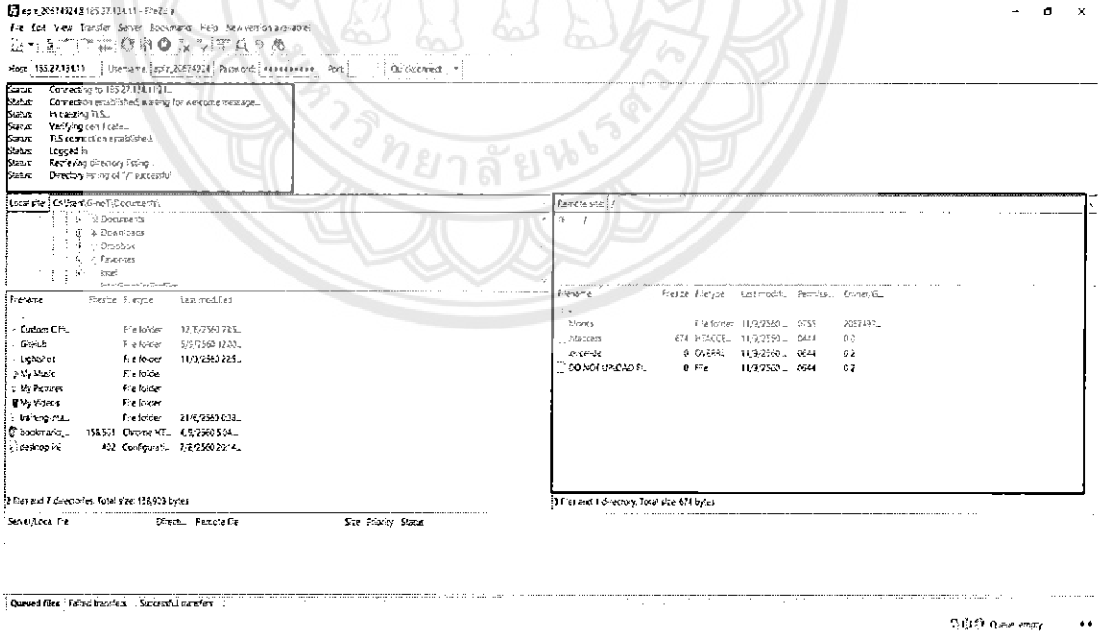
2. กรอก FTP Hostname, FTP Username, FTP Password และ FTP Port จากนั้นคลิกปุ่ม Quickconnect



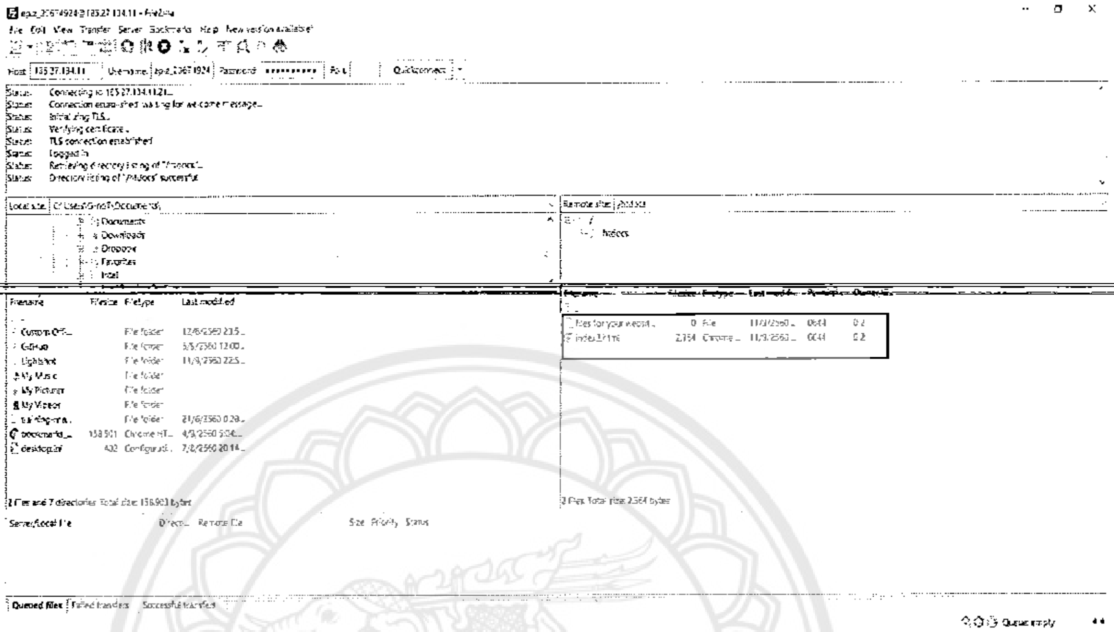
3. คลิกปุ่ม OK



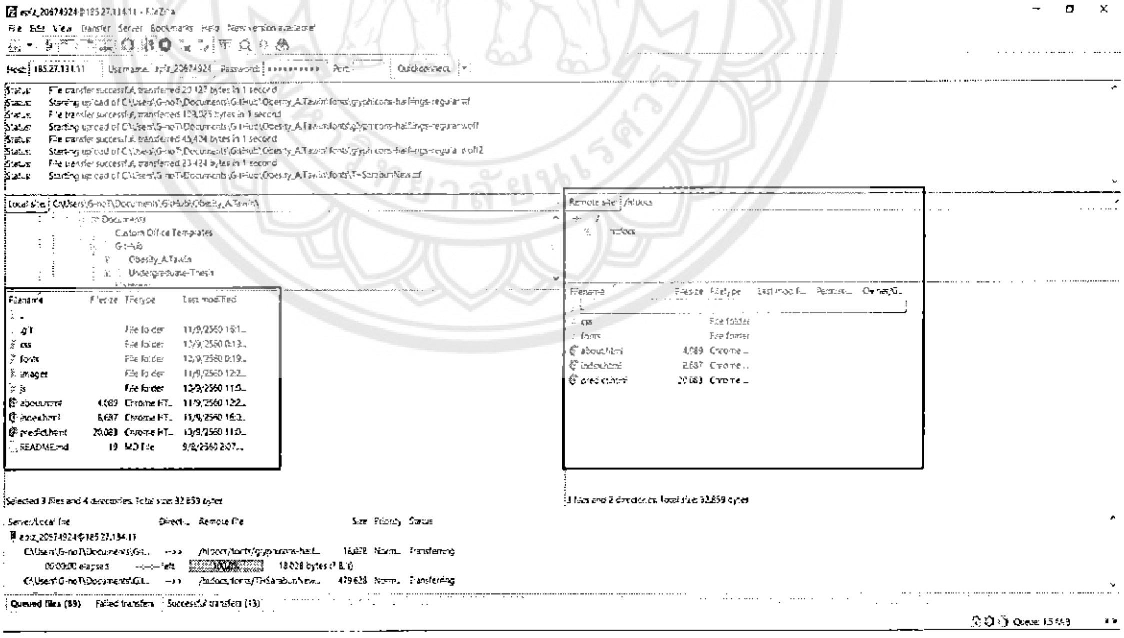
4. สถานะระบุว่าได้เข้าสู่ระบบเรียบร้อยแล้ว และด้านขวามือคือไฟล์ที่มีอยู่ใน Host : 158.27.134.11



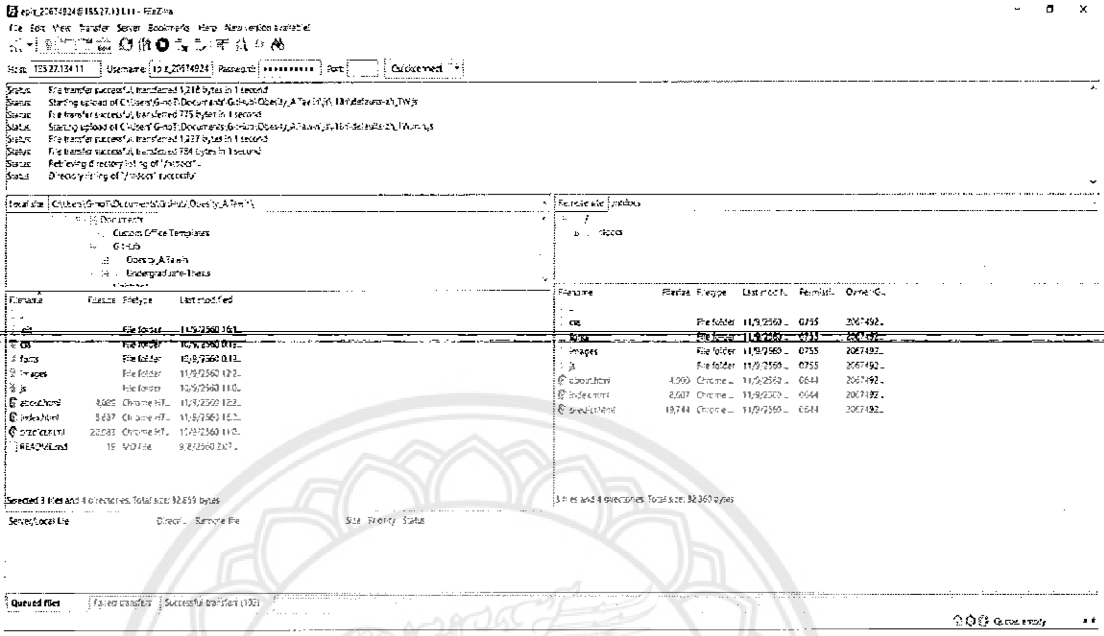
5. ลบไฟล์ใน /htdocs ทิ้ง



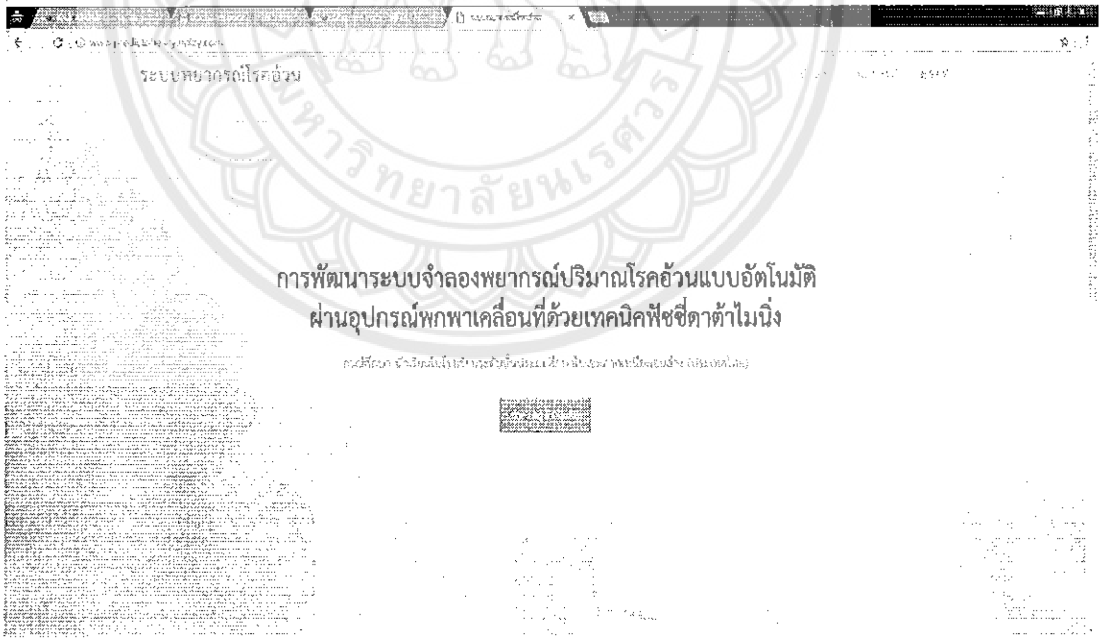
6. นำไฟล์ที่เราต้องการจากด้านซ้ายมือ มาใส่ในด้านขวามือ



7. เมื่อโปรแกรมอัพโหลดเสร็จสิ้น เราจะสามารถใช้งานเว็บไซต์ที่อัพโหลดเข้าไปได้

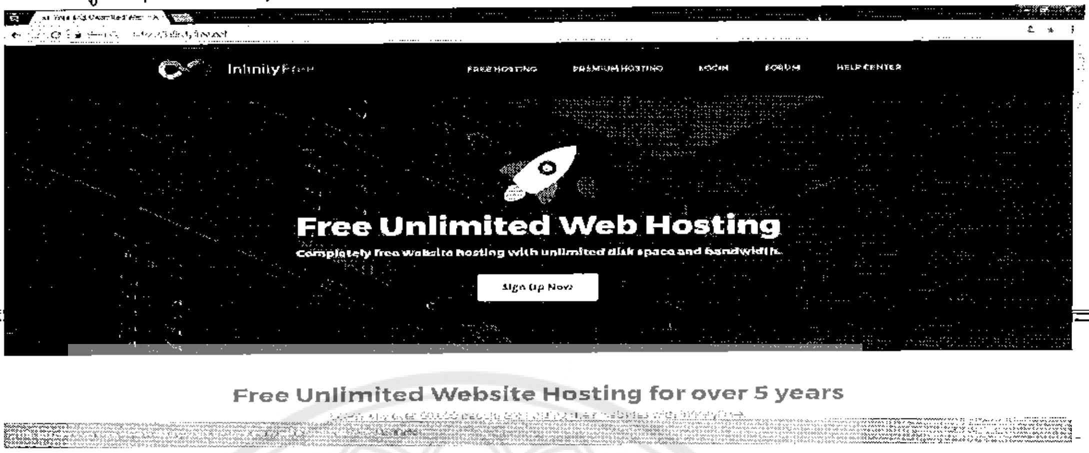


8. ให้เราเปิดไปที่ URL : www.predictobesity.epizy.com จะพบกับหน้าหลักของเว็บไซต์ที่เราทำการอัพโหลด



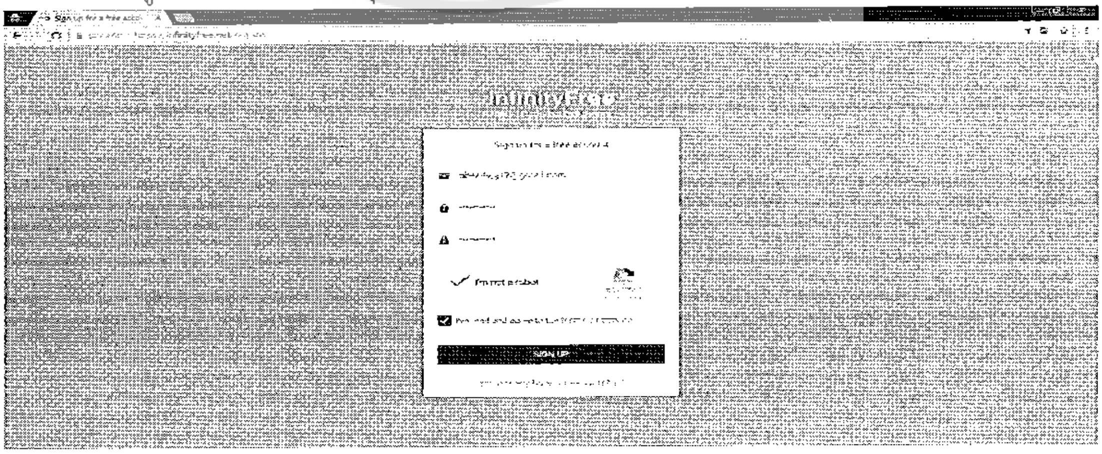
ภาคผนวก จ
การเข้าใช้งานบนโฮสของงานวิจัย

1. เข้าสู่ https://infinityfree.net/

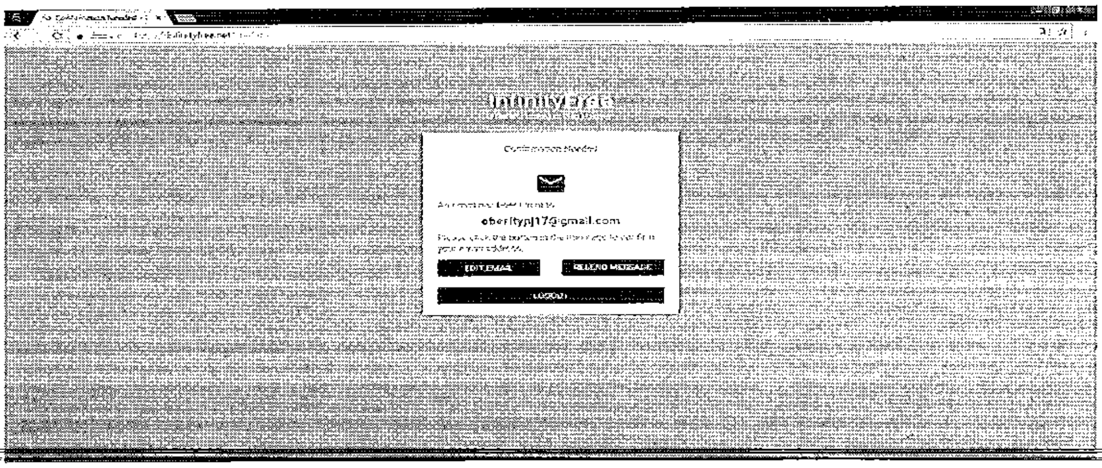


2. คลิกปุ่ม Sing up now

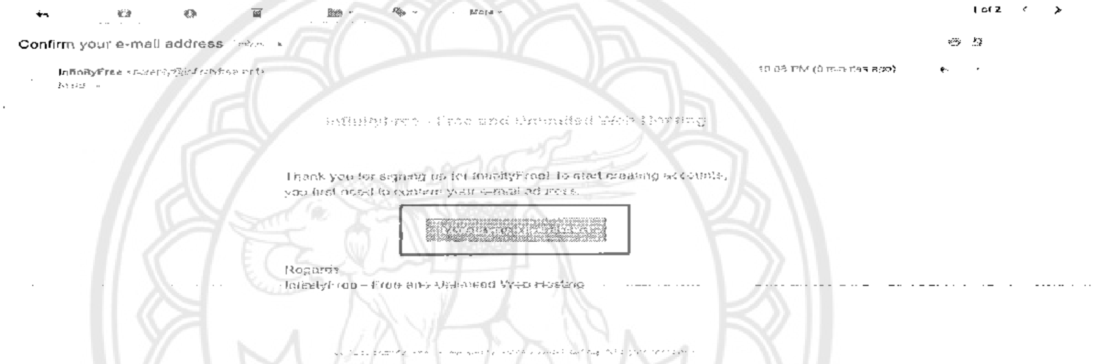
3. กรอกข้อมูลที่เกี่ยวข้องและกดปุ่ม SIGN UP



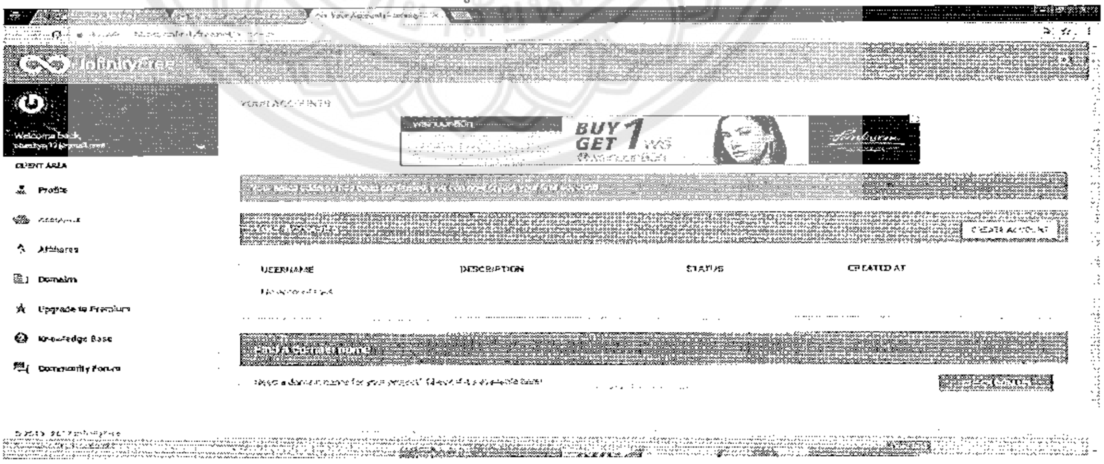
4. ระบบจะทำการส่ง mail ยืนยัน



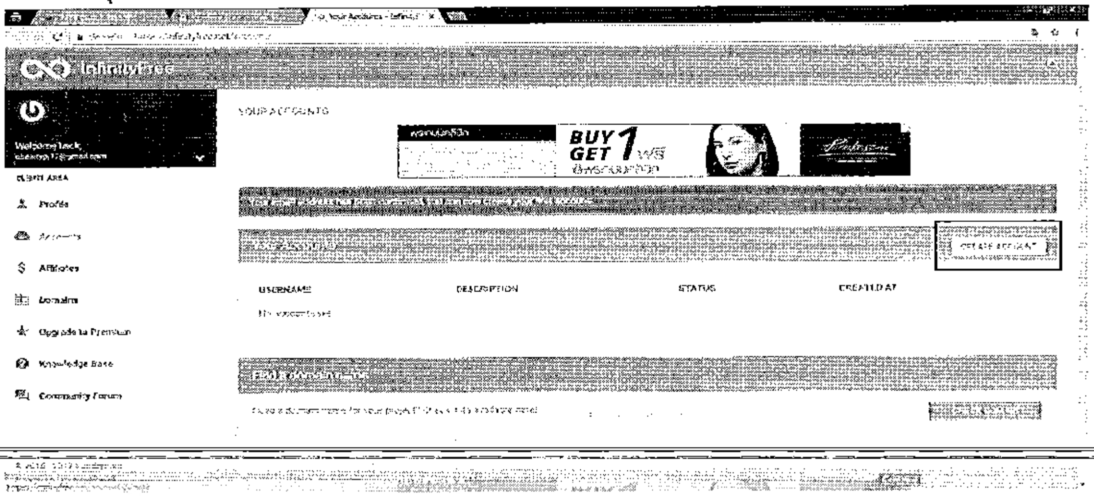
5. Login เข้าสู่ E-mail และทำการยืนยันตัวตน



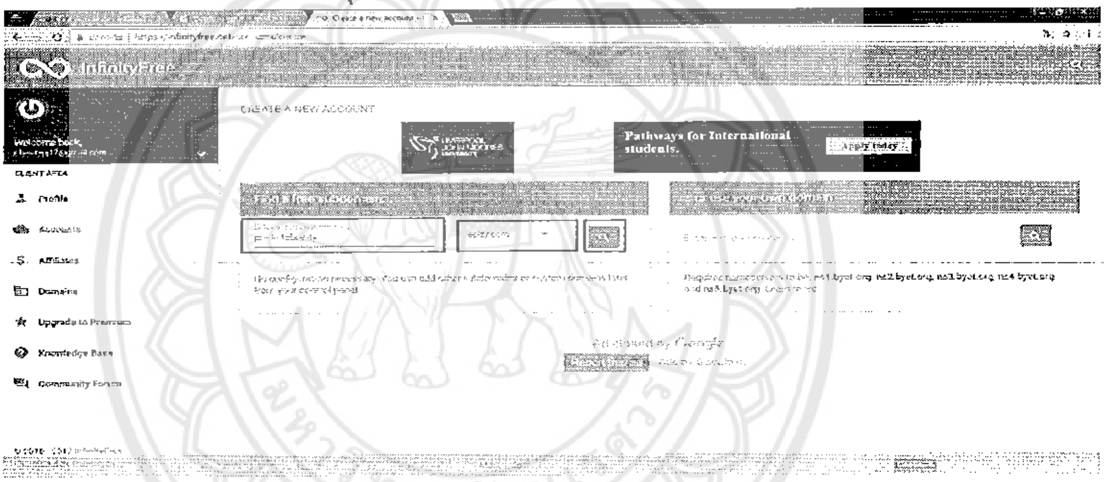
6. เมื่อยืนยันตัวตนเรียบร้อยแล้ว จะเข้ามาสู่หน้า Accounts



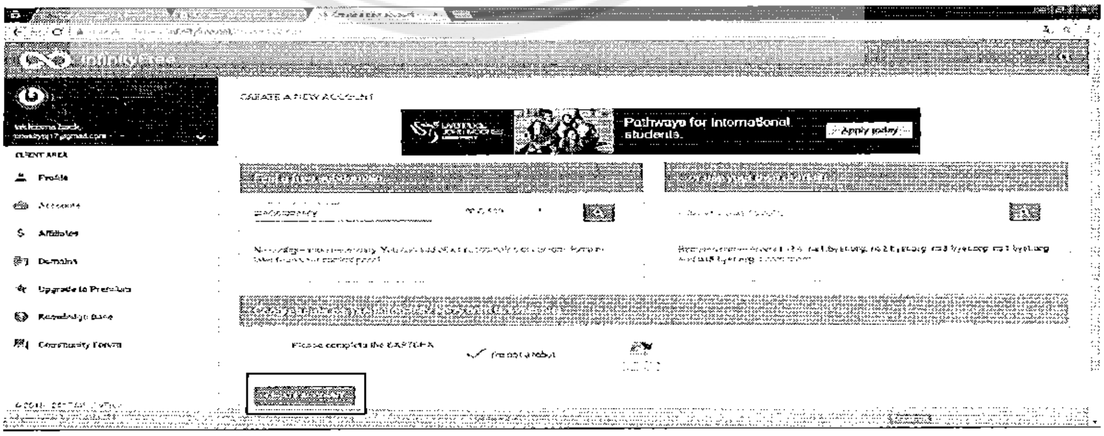
7. คลิกปุ่ม create account



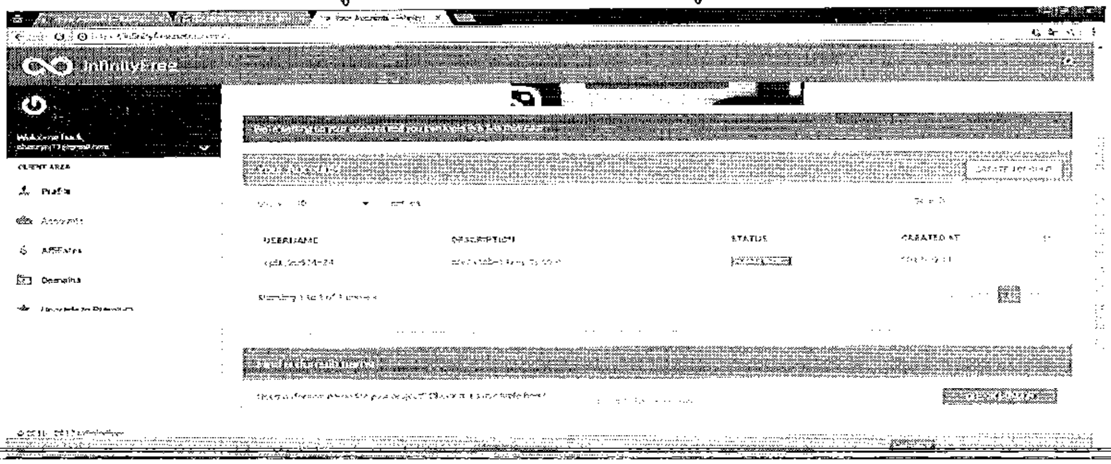
8. กรอก subdomain และกดปุ่มค้นหา



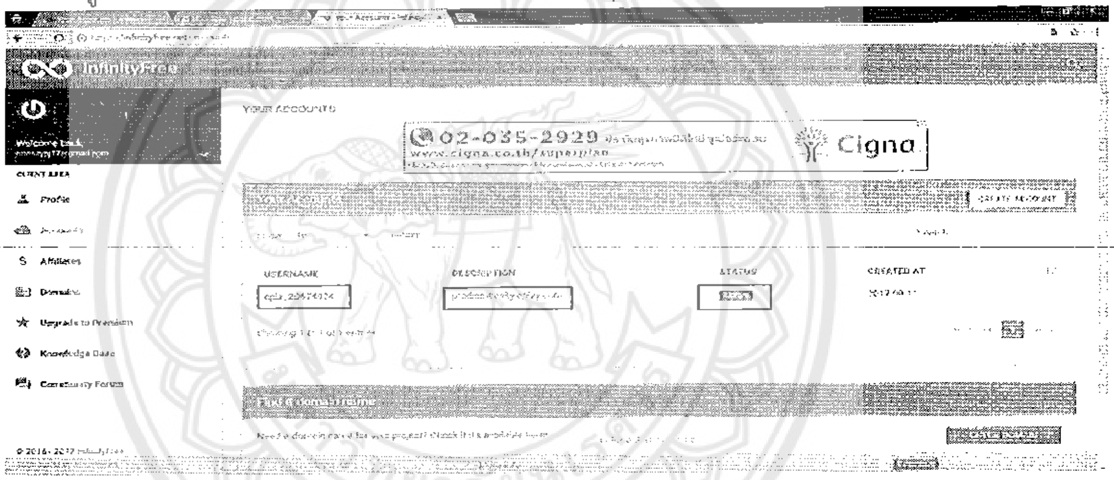
9. เมื่อตรวจสอบว่าไม่พบ subdomain ที่เหมือนกัน จะสามารถสร้าง subdomain ได้ด้วยการกดปุ่ม create account



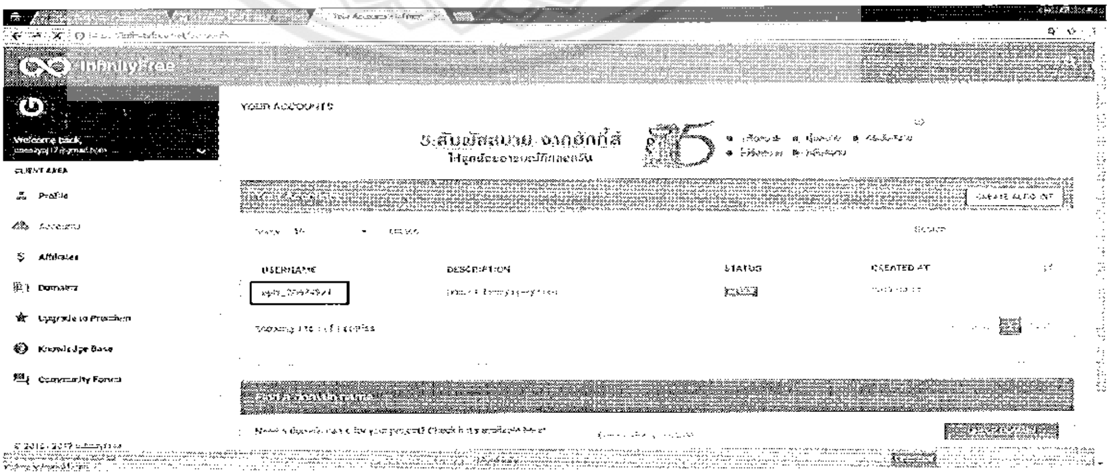
10. เมื่อสร้างเสร็จแล้ว จะนำมาสู่หน้า Accounts และใช้เวลาสักครู่ในการสร้าง



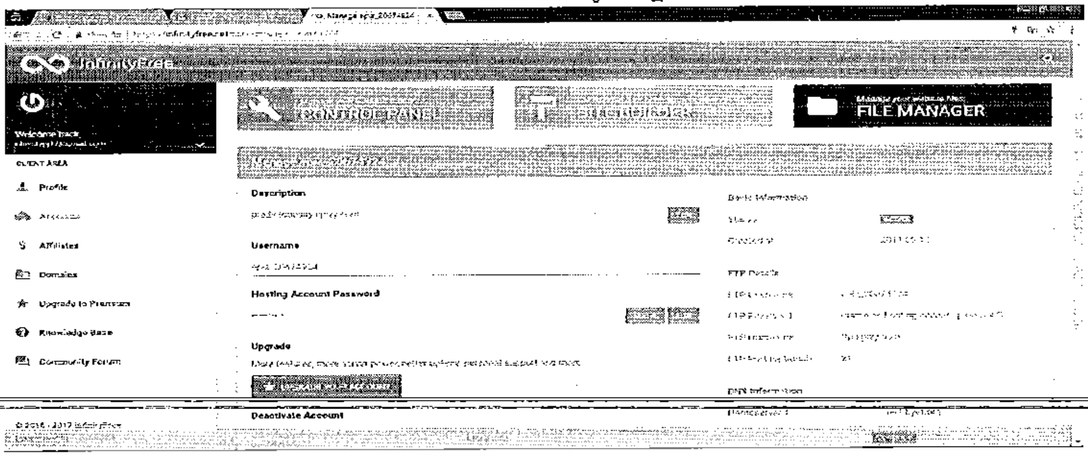
11. ในรูปจะเห็นว่าจะประกอบไปด้วย username description และ status หลังจากเราสร้าง



12. คลิกที่ username ที่เราต้องการอัพไฟล์เข้าไปบนเว็บ



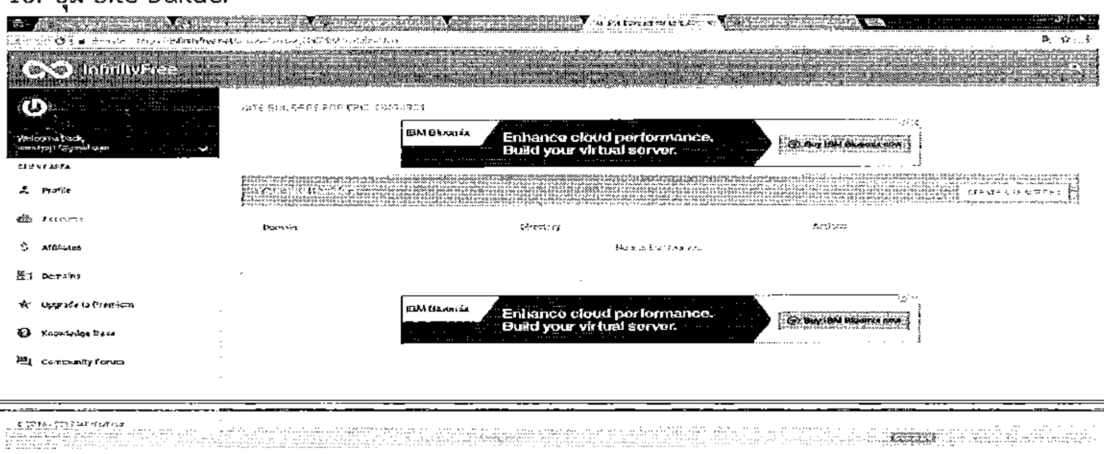
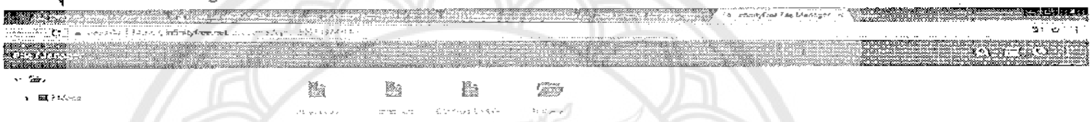
13. หลังจากที่เข้ามายัง username ที่ต้องการ จะพบข้อมูลพื้นฐานดังนี้



14. หากต้องการใช้งานอื่น สามารถใช้งานได้ผ่านปุ่มต่างๆดังนี้

15. ปุ่ม Control Panel

16. ปุ่ม Site Builder



17. ปุ่ม File Manager