การใช้ระบบปัญญาประดิษฐ์และกลุ่มข้อมูลราชภัฏจำแนกดัชนีมวลกายท้องถิ่น
USING RAJABHAT DATASET AND AI TO CLASSIFY COMMUNITY BMI

ไชย มีหนองหว้า[1] และ วุฒิชัย อินทร์แก้ว[2]
[1]สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์
[2]สาขาวิชาคอมพิวเตอร์ธุรกิจ มหาวิทยาลัยราชภัฏสงขลา
Chai Meenorngwar[1] and Wuttichai Inkaew[2]
[1]Information Technology, Valaya Alongkorn Rajabhat University under the Royal Patronage
[2]Business Computer, Songkhla Rajabhat University
(Received: January 26, 2023; Revised: August 22, 2023; Accepted: December 6, 2023)

บทคัดย่อ

มหาวิทยาลัยราชภัฏเป็นมหาวิทยาลัยที่มีความเข้าใจท้องถิ่นเป็นอย่างดี มีการพัฒนารูปแบบ ส่งเสริมให้ชุมชนในท้องถิ่นมีความเข้มแข็ง ผ่านทางกระบวนการ จัดการเรียนการสอน งานวิจัย และการบริการวิชาการ งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาการใช้ประโยชน์จากกลุ่มข้อมูลราชภัฏ (Rajabhat Dataset) ร่วมกับระบบปัญญาประดิษฐ์ (AI) เพื่อจำแนกและประเมินคุณภาพชีวิตของชุมชนในท้องถิ่น โดยใช้ดัชนีมวลกาย (BMI) ของประชากรเป็นตัวบ่งชี้ โดยใช้ข้อมูลจากจังหวัดปทุมธานีและสระแก้ว ซึ่งเป็นพื้นที่ให้บริการของมหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์ เป็นกรณีศึกษา มีการใช้รูปแบบการเรียนรู้ของเครื่อง (Machine Learning Model) เช่น Logistic Regression, SVM และ Decision Tree เพื่อทำนาย BMI ของประชากรในพื้นที่ ผลการวิจัยพบว่าโมเดล Decision Tree และ SVM มีความแม่นยำในการทำนายสูงสุด (0.997) งานวิจัยนี้แสดงให้เห็นถึงประโยชน์ของการนำกลุ่มข้อมูลราชภัฏมาใช้ร่วมกับระบบปัญญาประดิษฐ์ เพื่อกำหนดทิศทางในการดำเนินงานและการจัดกิจกรรมต่าง ๆ ของมหาวิทยาลัยราชภัฏได้อย่างมีประสิทธิภาพ โดยเฉพาะอย่างยิ่งในการส่งเสริมและพัฒนาคุณภาพชีวิตให้กับชุมชนในท้องถิ่น

คำสำคัญ: กลุ่มข้อมูลราชภัฏ, การเรียนรู้ของเครื่อง, ปัญญาประดิษฐ์, การจำแนกข้อมูล, ดัชนีมวลกาย

ABSTRACT

Rajabhat University is in an area-based university cluster that exchanges knowledge and technology between the university and the local community. It closely works with local populations by delivering courses, academic services, and research. A collective of 38 Rajabhat universities from around Thailand have developed a set of data called the "Rajabhat Dataset." It utilizes Valaya Alongkorn Rajabhat University (VRU) as the case study. VRU services both Pathum Thani and Sakaew provinces and contains 8,756 data samples tasked to learn/teach algorithms of the machine learning model with various approaches such as logistic regression, SVM/decision tree, and to predict the population's BMI in the areas. Lastly, AI predicts the QoL based on the Rajabhat data.

Keywords: Rajabhat Dataset, Machine Learning, Artificial Intelligence, Data Classification, Body Mass Index: BMI

บทนำ

การพัฒนารูปแบบของข้อมูล โดยเฉพาะอย่างยิ่งข้อมูลที่ใช้งานร่วมกันได้ ผ่านการออกแบบข้อมูล การจัดเก็บข้อมูล ในประเภทต่าง ๆ กระบวนการรวบรวมข้อมูลจากพื้นที่ ในการให้บริการของมหาวิทยาลัยราชภัฏ ในแต่ละแห่งด้วยมาตรฐานเดียวกัน จะกลายเป็น กลุ่มข้อมูล(Dataset) ที่จะแบ่งปัน(Shared) และใช้งานร่วมกัน (Reused) ตลอดจนการพัฒนาต่อยอดไปจนถึง ข้อมูลขนาดใหญ่ (Big Data) ได้ ด้วยเหตุผลดังกล่าว มหาวิทยาลัยราชภัฏทั้ง 38 แห่ง ได้ร่วมมือกันในการพัฒนากลุ่มข้อมูลราชภัฏ (Rajabhat Dataset) ขึ้น

ในงานวิจัยชิ้นนี้ ได้เล็งเห็นถึง การนำกลุ่มข้อมูลราชภัฏ ไปใช้งานโดยเฉพาะอย่างยิ่ง การจำแนกข้อมูลท้องถิ่น (Classification) และสามารถที่จะนำประโยชน์ของข้อมูล ไปใช้ในการพัฒนาท้องถิ่นได้อย่างมีประสิทธิภาพ อาทิ การพัฒนาคุณภาพชีวิตชุมชนในท้องถิ่น (Quality of Community Life) ซึ่งเป็นตัวชี้วัดที่สำคัญตามบริบทของมหาวิทยาลัยราชภัฏทุกแห่ง

ในบทความวิจัย (Saetang et al., 2021) ได้กล่าวถึงการยอมรับในการนำเอาเทคโนโลยีเข้ามาใช้งานจำเป็นจะต้อง ให้ผู้ใช้มองเห็นถึงประโยชน์ การประยุกต์เทคโนโลยีที่เกิดขึ้นเข้ากับการใช้งานในปัจจุบัน จาก อดีตจนถึงปัจจุบัน หน่วยงานของภาครัฐและเอกชน โดยเฉพาะอย่างยิ่ง มหาวิทยาลัยราชภัฏมีการจัดเก็บข้อมูลท้องถิ่นจำนวนมาก แต่ข้อมูลดังกล่าวอยู่ในรูปแบบของเอกสาร สื่อ สิ่งพิมพ์ เกิดความยากลำบากในการใช้งาน โดยเฉพาะการเชื่อมโยงข้อมูลระหว่างกัน จำเป็นต้องปรับเปลี่ยนข้อมูลให้เป็นรูปแบบของสื่อดิจิทัล (Digital Media) อาทิ ตารางข้อมูล หรือ ฐานข้อมูลต่าง ๆ ที่เกิดขึ้นในหน่วยงานต่าง ๆ นอกจากนี้การปรับรูปแบบและวิธีการเชื่อมต่อและการนำข้อมูลไปใช้งาน เพื่อข้อมูลสามารถที่จะใช้งานและเข้าถึงได้อย่างสะดวกและรวดเร็ว ที่สำคัญเรื่องของความปลอดภัย ในการเข้าถึงข้อมูล การปรับปรุงเปลี่ยนแปลงหรือแก้ไขข้อมูลต่าง ๆ สิทธิหรือการอนุญาตเข้าดำเนินการได้ เนื่องจากปริมาณของข้อมูลมีจำนวนเพิ่มมากขึ้นอย่างต่อเนื่อง การกับติดตามรายละเอียดของข้อมูลต่าง ๆ ผ่านทางกระดานติดตาม (Dashboard Monitoring) จะต้องถูกออกแบบมาให้สามารถที่จะรองรับความต้องการข้อมูลในมิติต่าง ๆ ที่จะเกิดขึ้นในปัจจุบันและอนาคตได้

บทบาทมหาวิทยาลัยราชภัฏ ในการเป็นมหาวิทยาลัยเพื่อการพัฒนาท้องถิ่น ที่มีความเด่นชัดและแตกต่างจากมหาวิทยาลัยแห่งอื่น ตลอดจนมีความได้เปรียบในการใกล้ชิดกับประชาชนในพื้นที่มาเป็นระยะเวลาที่ยาวนาน ตามพันธะกิจและวัตถุประสงค์ของการก่อตั้งมหาวิทยาลัยราชภัฏ เพื่อการสนับสนุนและถ่ายทอดองค์ความรู้ระหว่างชุมชนในท้องถิ่น การดำเนินงานตามพระบรมราโชบาย ในการถ่ายทอดองค์ความรู้หรือการศึกษา ลดช่องว่างและแก้ไขปัญหาความยากจน การสร้างความเข้มแข็งให้กับชุมชนในท้องถิ่น ให้สามารถสร้างรายได้และพึ่งตนเองได้อย่างมีประสิทธิภาพ ด้วยการพัฒนาท้องถิ่น ผ่านทางศาสตร์ทางด้านการศึกษา นำองค์ความรู้ต่าง ๆ ที่มีอยู่ในท้องถิ่นมาพัฒนาและสร้างรายได้ บริหารการจัดการตนเอง ทั้งในชุมชน ระหว่างชุมชนและชุมชนภายนอก ได้อย่างยั่งยืน

ทั้งนี้งานวิจัยชิ้นนี้เป็นตัวอย่างของการพบปัญหาของการบริหารจัดการข้อมูล โดยเฉพาะอย่างยิ่ง การใช้ข้อมูลระหว่างหน่วยงานต่าง ๆ เป็นอุปสรรคในการเชื่อมหรือแลกเปลี่ยนข้อมูลระหว่างมหาวิทยาลัยราชภัฏ ตลอดจนข้อมูลมีรูปแบบที่หลากหลายและมีความแตกต่างกัน ดังนั้นการพัฒนามาตรฐานรูปแบบข้อมูลภายใต้ชื่อฐานข้อมูลหรือกลุ่มข้อมูลราชภัฏ “Rajabhat Dataset” จะช่วยให้ เกิดประโยชน์ ในการสร้างความยั่งยืนทางด้านเทคโนโลยีสารสนเทศของข้อมูล ด้วยมาตรฐานข้อมูลเดียวกัน ตลอดจนการนำศาสตร์ทางด้านวิทยาการข้อมูล (Data Science) เข้ามาประยุกต์ใช้ นำเสนอแนวโน้มของสถานการณ์หรือคาดการณ์ผลลัพธ์ที่อาจจะเกิดขึ้นในอนาคต จากการจัดเก็บรวบรวมข้อมูลอย่างต่อเนื่อง จะสามารถพัฒนาเป็นฐานข้อมูลขนาดใหญ่ (Big Data) เพื่อประมวลผลในรูปแบบของข้อมูลขนาดใหญ่ระหว่างมหาวิทยาลัยราชภัฏได้

ข้อมูลที่จัดรูปแบบแล้วยังรองรับการนำระบบปัญญาประดิษฐ์มาสนับสนุนการใช้งาน ตัวอย่างงานวิจัย (Sulochana et al., 2022) ที่แสดงให้เห็นว่าใช้แมชชีนเลิร์นนิง (Machine Learning) ในการทำนายรูปแบบการจ้างงานและการเข้าร่วมโปรแกรมกลางวันสำหรับผู้ที่มีความบกพร่องทางสติปัญญาและพัฒนาการ (IDD) โดยมีเป้าหมายเพื่อช่วยในการวางแผนและจัดสรรทรัพยากรให้มีประสิทธิภาพมากขึ้น โดยแมชชีนเลิร์นนิงสามารถนำมาใช้ในการทำนายรูปแบบการจ้างงานและการเข้าร่วมโปรแกรมกลางวันสำหรับผู้ที่มี IDD ได้อย่างมีประสิทธิภาพ ซึ่งสามารถเป็นประโยชน์ในการวางแผนและจัดสรรทรัพยากรเพื่อสนับสนุนผู้ที่มี IDD ให้มีคุณภาพชีวิตที่ดีขึ้น

เพื่อให้เกิดความชัดเจน งานวิจัยชิ้นนี้ ได้นำเสนอวิธีการวิเคราะห์คุณภาพชีวิตของท้องถิ่น (Quality of Community Life) ด้วยดัชนีมวลกาย (Body Mass Index) เพื่อเป็นอย่างในการวิเคราะห์ตัวที่บ่งชี้ความเป็นอยู่และคุณภาพชีวิตของชุมชน ตามที่องค์การอนามัยโลก (World Health Organization: WHO) ได้นิยามถึงเรื่องของคุณภาพชีวิต ที่สามารถแบ่งออกเป็น 4 ด้านด้วยกัน คือ ด้านร่างกาย ด้านจิตใจ ด้านสังคม และด้านสิ่งแวดล้อม การนำดัชนีมวลรวม ผลลัพธ์ที่ได้ยังใช้เป็นตัวติดตามความก้าวหน้าของโครงการแต่ละตำบลในท้องถิ่นที่มหาวิทยาลัยราชภัฏได้ลงไปสนับสนุนและส่งเสริมในแต่ละพื้นที่ของมหาวิทยาลัยราชภัฏที่รับผิดชอบ

มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์ รับผิดชอบชุมชนในท้องถิ่น ในพื้นที่จังหวัดปทุมธานีและสระแก้ว และได้ดำเนินการโครงการวิจัย โครงการส่งเสริมการเรียนรู้และบริการวิชาการ ตลอดจนโครงการส่งเสริมทำนุบำรุงศิลปวัฒนธรรม ผ่านความร่วมมือระหว่าง คณาจารย์ บุคลากรและนักเรียนนักศึกษาจากคณะและหลักสูตรต่าง ๆ กับชุมชนในท้องถิ่นมาอย่างต่อเนื่อง เรียกได้ว่าท้องถิ่นคือ ตลาดแรงงาน (Labor Market หรือ Marketplace) ที่สำคัญของมหาวิทยาลัยราชภัฏ มีความเข้าใจ ถึงบริบทของชุมชนในท้องถิ่นมากกว่าสถาบันการศึกษาอื่น เห็นได้จากความร่วมมือจากโครงการวิจัย หรือโครงการบริการวิชาการจำนวนมาก ที่มีการรวบรวมข้อมูลชุมชน ภูมิปัญญาท้องถิ่นจากพื้นที่ต่าง ๆ เพื่อใช้ในการติดตามความสำเร็จและประสิทธิภาพของการดำเนินงาน ในแต่ละปีงบประมาณ อย่างไรก็ตาม ปัญหาของ การกำกับติดตามข้อมูลท้องถิ่นที่จัดเก็บรวบรวม ข้อมูลอยู่กระจายและการเข้าถึงข้อมูลในแต่ละแหล่งข้อมูล ไม่สามารถทำได้สะดวก

จากปริมาณข้อมูลที่จัดเก็บรวบรวมมีปริมาณมาก ต้องมีวิธีการจำแนกข้อมูล เพื่อใช้ข้อมูลมาสนับสนุนการตัดสินใจในมิติต่าง ๆ อาทิ ตัวอย่างงานวิจัยของ (Zhao et al., 2022) ได้ออกแบบการทำนายลูกค้าที่จะยกเลิกบริการ (customer churn) ถือเป็นความท้าทายสำคัญสำหรับธุรกิจต่าง ๆ บทความวิจัยนี้ได้นำเสนอการศึกษาเปรียบเทียบการใช้โมเดล Decision Tree และ Random Forest ในการแก้ปัญหานี้ โดยใช้ข้อมูลการขายจากบริษัทเคมีแห่งหนึ่งตั้งแต่ปี 2012 ถึง 2020 โดยพิจารณาจาก ปัจจัยที่มีผลต่อการยกเลิกบริการ จำนวนครั้งที่ซื้อสินค้าราคาต่ำ (Low-priced count: LC) ยอดเงินรวมที่ใช้จ่าย (Total amount of money: TM) ระยะเวลาตั้งแต่เริ่มเป็นลูกค้า (Creation time: CT) ผลการวิจัยพบว่า จำนวนครั้งที่ซื้อสินค้าราคาต่ำ (LC) เป็นปัจจัยที่สำคัญที่สุดในการทำนายลูกค้าที่จะยกเลิกบริการ และมีการเปรียบเทียบประสิทธิภาพของโมเดล Random Forest ให้ผลการทำนายที่แม่นยำกว่า Decision Tree เมื่อพิจารณาจากค่าต่าง ๆ เช่น Confusion Matrix, ROC Curve, AUC, Precision, Recall และ F1 Score สำหรับการทำนายลูกค้าที่จะยกเลิกบริการล่วงหน้าช่วยให้ธุรกิจสามารถวางแผนและดำเนินกลยุทธ์เพื่อรักษาลูกค้าไว้ได้ การใช้ Machine Learning โดยเฉพาะโมเดล Random Forest เป็นเครื่องมือที่มีประสิทธิภาพในการแก้ปัญหานี้

ดังนั้น งานวิจัยการพัฒนากลุ่มข้อมูลราชภัฏ เป็นส่วนหนึ่งของการแก้ไขปัญหาการใช้ข้อมูลเพื่อเสริมสร้างความเข้มแข็งและคุณภาพชีวิตของท้องถิ่น ในการประมวลผลจากกลุ่มข้อมูลราชภัฏ จะใช้หลักการเรียนรู้ของเครื่อง (Machine Learning) ตามกระบวนการระบบปัญญาประดิษฐ์ (Artificial Intelligence) เพื่อ

สามารถที่จะวิเคราะห์ความสัมพันธ์และปัจจัยหรือคุณลักษณะ (Feature) ที่มีผลต่อข้อมูลหรือผลการคาดการณ์ (Class) ตามหลักการวิทยาการข้อมูล ในการแบ่งแยกกลุ่มของข้อมูล (Classification) ในกลุ่มต่าง ๆ เพื่อเรียนรู้รูปแบบข้อมูล (Data Pattern) อาทิเช่น ดัชนีมวลกายที่คำนวณจากน้ำหนักและส่วนสูงของชุมชน เพื่อเปรียบเทียบคุณภาพชีวิตของประชากรในท้องถิ่น เป็นตัวอย่างที่ศึกษารูปแบบการเข้าถึงข้อมูลต่าง ๆ และวิธีการรวบรวมข้อมูลในพื้นที่มานำเสนอข้อมูลในรูปแบบของรายงานกระดานติดตาม (Dashboard) ประกอบด้วย ภาพแผนภูมิ และค่าคุณภาพชีวิตของท้องถิ่นที่อยู่ในระดับดีขึ้น (Well-Being) ผลของการนำเสนอในรูปแบบนี้ จะช่วยให้มหาวิทยาลัยราชภัฏทราบแนวทางในการพัฒนาจุดเด่นและส่งเสริมจุดแข็งของชุมชนในท้องถิ่นได้มากยิ่งขึ้น อีกทั้ง เป็นการสร้างความเข้าใจและชัดเจนระหว่างมหาวิทยาลัยกับท้องถิ่น นับเป็นการประหยัดเวลาและงบประมาณในการตอบสนองต่อความต้องการ ได้ตรงเป้าหมายและทันท่วงที

วัตถุประสงค์การวิจัย

งานวิจัยชิ้นนี้ ต้องการที่จะศึกษาถึงวิธีการในการ ออกแบบและพัฒนามาตรฐานของการจัดเก็บข้อมูลท้องถิ่นให้อยู่ในรูปแบบเดียวกัน และสามารถที่จะใช้ในรูปแบบของข้อมูลในรูปแบบเปิด(Open Data) ให้นักวิจัย นักวิชาการ หรือบุคคลทั่วไป สามารถที่จะนำข้อมูลดังกล่าวไปใช้ประโยชน์ได้มากขึ้น ดังนั้นวัตถุประสงค์ของงานวิจัยชิ้นนี้จึงมีดังต่อไปนี้

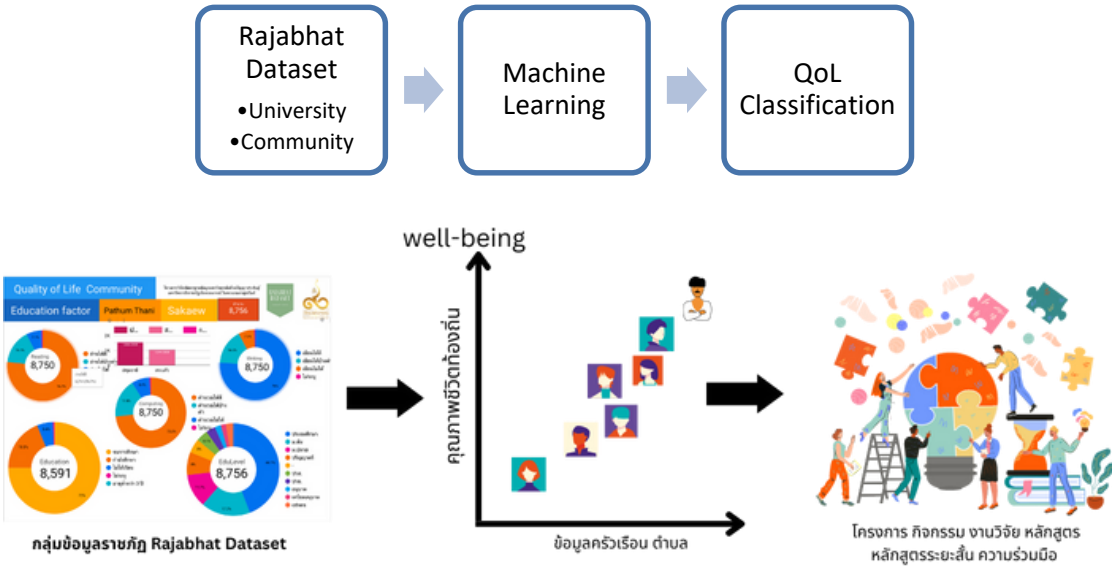
1. เพื่อวิเคราะห์และทำนายดัชนีมวลกายของประชากรในท้องถิ่นนำเสนอ เป็นตัวอย่างของการใช้ระบบปัญญาประดิษฐ์และกลุ่มข้อมูลราชภัฏ
2. เพื่อจำแนกคุณภาพชีวิตของชุมชนในแต่ละท้องถิ่นจากดัชนีมวลกายได้

วิธีดำเนินการวิจัย

1. ประชากรและกลุ่มตัวอย่าง

ในงานวิจัยชิ้นนี้ได้นำกลุ่มข้อมูลท้องถิ่น ซึ่งเป็นหนึ่งในกลุ่มข้อมูลราชภัฏที่มีการออกแบบและจัดเก็บจากข้อมูลในพื้นที่ให้บริการของมหาวิทยาลัยราชภัฏ นำระบบปัญญาประดิษฐ์มาวิเคราะห์และพัฒนาการเรียนรู้ของเครื่อง เพื่อจำแนกลักษณะข้อมูลในการบริหารและสนับสนุนกิจกรรมโครงการต่าง ๆ ที่ถ่ายทอดลงไปยังท้องถิ่น ผลลัพธ์ที่เกิดขึ้นเป็นการแลกเปลี่ยนเรียนรู้ ระหว่างท้องถิ่นกับทางมหาวิทยาลัยราชภัฏได้อย่างมีประสิทธิภาพ

การเริ่มต้นจัดเก็บข้อมูลจากประชากร ใช้ฐานข้อมูลจำนวน 2 ฐานข้อมูล คือ ฐานข้อมูลมหาวิทยาลัย (University) และฐานข้อมูลท้องถิ่น (Community) เพื่อจัดเก็บตัวอย่างประชากรด้วยแบบสอบถามในการลงพื้นที่จัดเก็บข้อมูลของมหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์ ในตำบลต่าง ๆ ของจังหวัดปทุมธานีและสระแก้ว ซึ่งรูปแบบของการจัดเก็บข้อมูลนี้จะมีวิธีการในการเก็บรวบรวมข้อมูลผ่านทางเครือข่าย องค์การบริหารส่วนตำบล และอาสาสมัครสาธารณสุข (อ.ส.ม.) ของโรงพยาบาลส่งเสริมสุขภาพตำบล (รพ.สต.) จากงานวิจัยพบว่า เป็นรูปแบบที่มีความน่าเชื่อถือทางด้านข้อมูล เนื่องจากสมาชิก อสม.มีความใกล้ชิดกับชุมชนในท้องถิ่นในแต่ละครัวเรือนได้เป็นอย่างดี



ภาพที่ 1 กระบวนการ จำแนกคุณภาพชีวิตของท้องถิ่นด้วย Machine : AI connect with Rajabhat Dataset

จากภาพที่ 1 ข้อมูลครัวเรือนจากกลุ่มข้อมูลราชภัฏ ประกอบด้วย ตารางข้อมูล(Data Table) และฐานข้อมูล (Database) เลือกประเภทกลุ่มข้อมูล (Dataset) เพื่อประมวลผล และแสดงข้อมูลที่สนใจ ด้วยแผนภูมิชนิดต่าง ๆ ปรับเปลี่ยนข้อมูลครัวเรือนในระดับตำบลเพื่อใช้ในการประมวลเพื่อหาตัวชี้วัดคุณภาพชีวิตด้วยการหาดัชนีมวลกาย (BMI) ของประชากรในท้องถิ่น จากข้อมูลส่วนสูงและน้ำหนัก ตลอดจนการเปรียบเทียบช่วงอายุของประชากรในแต่ละตำบล และส่งต่อไปให้อัลกอริทึมตามหลักการเรียนรู้ของเครื่อง (Machine Learning algorithms) เป็นข้อมูลที่ใช้เรียนรู้และจำแนกความแตกต่างของข้อมูล เรียกว่า “รูปแบบของข้อมูล” (Data Pattern) ที่เกิดขึ้น ท้ายที่สุด ผลของการประมวลผลค่า ดัชนีมวลกาย (BMI) จะใช้ในการออกแบบ กิจกรรม งานวิจัย และหลักสูตร ต่าง ๆ ภายใต้ความร่วมมือระหว่างมหาวิทยาลัยกับท้องถิ่นนั่นเอง

2. เครื่องมือที่ใช้ในการวิจัย

ฐานข้อมูล (Database) ถือว่าเป็นเครื่องมือและหัวใจสำคัญ ในการที่เราจะนำไปใช้ในการสนับสนุนและแก้ไขตลอดจนการวิเคราะห์ปัญหาที่เกี่ยวข้อง ในงานวิจัยชิ้นนี้ มีเป้าหมายในการที่จะตอบสนองต่อความต้องการของชุมชนในท้องถิ่น โดยอาศัยศักยภาพและข้อมูลที่มีอยู่ของทางมหาวิทยาลัยราชภัฏ ดังนั้น ข้อมูลที่สำคัญที่ใช้เป็น พื้นฐานในการวิเคราะห์ อาทิเช่น ข้อมูลทางด้านองค์ความรู้ จากปราชน์หรือบุคคลากรที่มีความรู้ความชำนาญที่เกี่ยวข้อง ตลอดจนองค์ความรู้ที่มีอยู่ในท้องถิ่น จะนำมาจัดเก็บเป็นฐานข้อมูลเพื่อใช้ในการประสานและเชื่อมโยงกับศักยภาพของท้องถิ่น เพื่อใช้ในการแก้ไขปัญหาต่าง ๆ ขนสมบัติของฐานข้อมูลก็คือ เข้าถึงหรือค้นหาได้ตลอดเวลา ผ่านทางเครือข่ายอินเทอร์เน็ต และอุปกรณ์อิเล็กทรอนิกส์ อาทิ เครื่องคอมพิวเตอร์ โทรศัพท์ไอแพด เป็นต้น

ในงานวิจัยนี้อธิบายวิธีการใช้ข้อมูล สามารถจะเริ่มจากการตั้ง “คำถาม” ที่ถือว่าเป็นความต้องการที่ในการนำจากข้อมูลท้องถิ่น และเป็นส่วนสำคัญของกระบวนการวิเคราะห์ข้อมูล เพื่อหารูปแบบของการเรียนรู้ให้เครื่องหรือที่เราเรียกกันว่า “Learning” เป็นส่วนหนึ่งของกระบวนการระบบปัญญาประดิษฐ์ (Artificial Intelligence: AI) ด้วยการอาศัยการวิเคราะห์ จากข้อมูลที่เกิดขึ้นในอดีต ดังนั้นข้อมูลที่มีการจัดเก็บ จากท้องถิ่น

ในรอบแรกจะถูกใช้เป็นข้อมูลเพื่อทำการทดสอบโมเดล ที่ถูกพัฒนาขึ้น เพื่อหาคำตอบของ ความน่าเชื่อถือและความถูกต้องของข้อมูล ปัญหาของการ กรอกหรือบันทึกของข้อมูลจะถูกปรับปรุงในขั้นตอนนี้ เพื่อที่จะใช้ในการปรับปรุงคุณภาพของข้อมูลในรอบถัดไป ในงานวิจัยนี้ นำเสนอรูปแบบผลการรวบรวมข้อมูลเพื่อใช้ติดตามข้อมูล หรือที่เราเรียกกันว่า “แดรชบอร์ด” หรือ “Dashbaord” เพื่อความสะดวก และง่ายในการใช้งาน

กระดานติดตามข้อมูล (Dashboard) ถือว่าเป็นเครื่องมือที่มีประโยชน์ ในการสร้างความเข้าใจเกี่ยวกับข้อมูล ให้ชัดเจนมากยิ่งขึ้น รายละเอียดมิติของการติดตาม รูปแบบที่หลากหลาย ทั้งในรูปแบบจากข้อมูลโดยตรง และผ่านกระบวนการ คำนวณประมวลผลทางสถิติ เพื่อสร้างการเชื่อมโยง ข้อมูลระหว่างกัน ในมิติต่าง ๆ ทั้งในเรื่องของพื้นที่ ประเภทของข้อมูล และกลุ่มเป้าหมายที่ต้องการ ได้สะดวกและรวดเร็ว

ตัวอย่าง การใช้กลุ่มข้อมูลราชภัฏ คำนวณหา BMI เพื่อเป็นตัวชี้วัดคุณภาพชีวิตด้านสุขภาพของชุมชนในท้องถิ่น

ประชากร ชุมชนที่อาศัยอยู่ใน 4 อำเภอ 8 ตำบล กลุ่มเป้าหมายของจังหวัดสระแก้วและปทุมธานี

วิธีการจัดเก็บกลุ่มตัวอย่าง

ข้อมูลท้องถิ่นที่กลุ่มข้อมูลราชภัฏมีการจัดเก็บ จะแบ่งออกเป็น 9 มิติข้อมูล ได้แก่ ข้อมูลพื้นที่จังหวัด ข้อมูลด้านที่ตั้ง ข้อมูลด้านพื้นฐานครัวเรือน ข้อมูลด้านสุขภาพครัวเรือน ข้อมูลด้านเศรษฐกิจรายรับรายจ่าย ภาวะหนี้สิน ข้อมูลด้านสิ่งแวดล้อม ข้อมูลด้านการเมืองการปกครอง และข้อมูลด้านช่องทางการสื่อสาร ข้อมูลทางด้านครัวเรือน จะเป็นข้อมูลหลักที่ใช้สำหรับในการวิเคราะห์ค่าดัชนีกระบวนการต่าง ๆ โดยใช้ค่าส่วนสูงและน้ำหนักตลอดจนอายุของ ประชากรที่อยู่ในชุมชนในตำบล เพื่อหาค่า ดัชนีมวลกาย (Body Mass Index) เพื่อใช้ทำนายคุณภาพชีวิตของชุมชนในแต่ละพื้นที่ในอนาคต

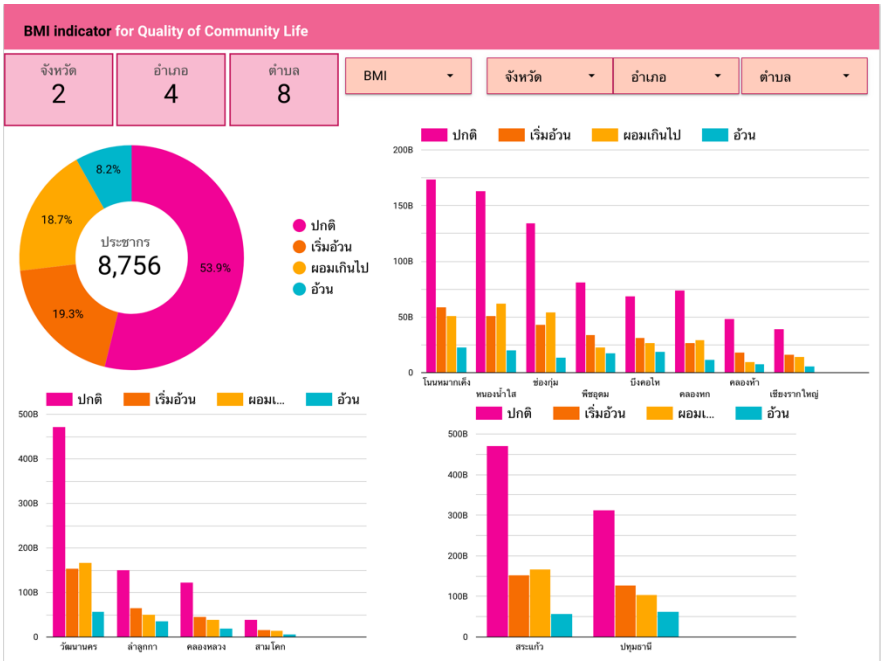
กระบวนการวิเคราะห์

จากภาพที่ 2 จะเป็นการนำ กลุ่มข้อมูลราชภัฏ ข้อมูลชุมชนในท้องถิ่น เฉพาะด้านข้อมูลพื้นฐานของประชากรในครัวเรือน ของจังหวัดปทุมธานีและสระแก้ว 4 อำเภอ 8 ตำบลเป็นตัวอย่างของการวิเคราะห์ โดยการหาข้อมูลหลัก (Feature) ที่ เหมาะสมสำหรับระบบปัญญาประดิษฐ์ (AI) ในการวิเคราะห์จะประกอบด้วยส่วนสูงและน้ำหนักของประชากรที่อยู่ในตำบลกลุ่มตัวอย่าง มาทำการคำนวณหาค่า BMI เพื่อเป็นตัวอย่าง

ปัจจัยตัวชี้วัดคุณภาพชีวิตของแต่ละชุมชน มีหลายวิธีด้วยกัน ในการหาค่าดัชนีมวลกายเป็นเพียงวิธีหนึ่งเท่านั้น เพื่อเป็นตัวอย่างในการกำกับติดตาม การพัฒนาคุณภาพชีวิตให้กับชุมชนในท้องถิ่นอย่างต่อเนื่อง มหาวิทยาลัยราชภัฏจะนำข้อมูลดังกล่าว ออกแบบหลักสูตรระยะสั้น หรือกิจกรรมที่เหมาะสมเพื่อแก้ไขปัญหาหรือสนับสนุนชุมชนในท้องถิ่น ได้อย่างมีประสิทธิภาพ ทันต่อการปรับเปลี่ยนของสภาพแวดล้อม และเทคโนโลยีที่เกิดขึ้นอย่างรวดเร็ว ปัญญาประดิษฐ์จะคำนวณ คุณภาพชีวิต หรือคาดการณ์ค่าของดัชนีมวลกายของชุมชนที่เกิดขึ้นในอนาคต นำความรู้หรือผลการคาดการณ์ที่ได้รับในปัจจุบันหรือในอดีต เป็นฐานข้อมูลที่สำคัญ ในการคำนวณหาค่าดัชนีมวลกายที่เกิดขึ้น ของท้องถิ่นต่าง ๆ ตามระยะเวลาที่กำหนด

คุณภาพชีวิตของชุมชนต้องมีการออกแบบกิจกรรมส่งเสริม ป้องกันเพื่อให้มีสุขภาพที่ดีขึ้น ตัวอย่างงานวิจัย(Tamayo et al., 2021) ที่มีกิจกรรมต่างเพื่อเข้าไปทำการแทรกแซง(Intervention) ในการศึกษาเกี่ยวกับการควบคุมโรคอ้วนในเด็กเชื้อสายฮิสแปนิกในสหรัฐฯ ผ่านการแทรกแซงที่เน้นครอบครัว โดยพบว่าการแทรกแซงส่วนใหญ่ที่มีอยู่ในปัจจุบันมีประสิทธิภาพในการควบคุมโรคอ้วนในเด็กฮิสแปนิก การแทรกแซงระยะสั้น (8-36 สัปดาห์) พบการเปลี่ยนแปลงในพฤติกรรมสุขภาพ (เช่น การบริโภคเครื่องดื่มที่มีน้ำตาล, เวลา

อยู่หน้าจอ) และผลลัพธ์ด้านสุขภาพ (เช่น คุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ) แต่หลายคนไม่เห็นการเปลี่ยนแปลงในตัวแปรทางมานุษยวิทยา (เช่น ดัชนีมวลกาย [BMI], ความดันโลหิต) การแทรกแซงที่วัดในระยะยาว (เช่น 48-144 สัปดาห์) พบว่า มีการลดลงของพฤติกรรมที่ยั่งยืน (เช่น ปริมาณแคลอรี่ที่บริโภค) และมาตรการทางมานุษยวิทยา



ภาพที่ 2 ข้อมูล BMI คุณภาพชีวิตของชุมชนท้องถิ่น

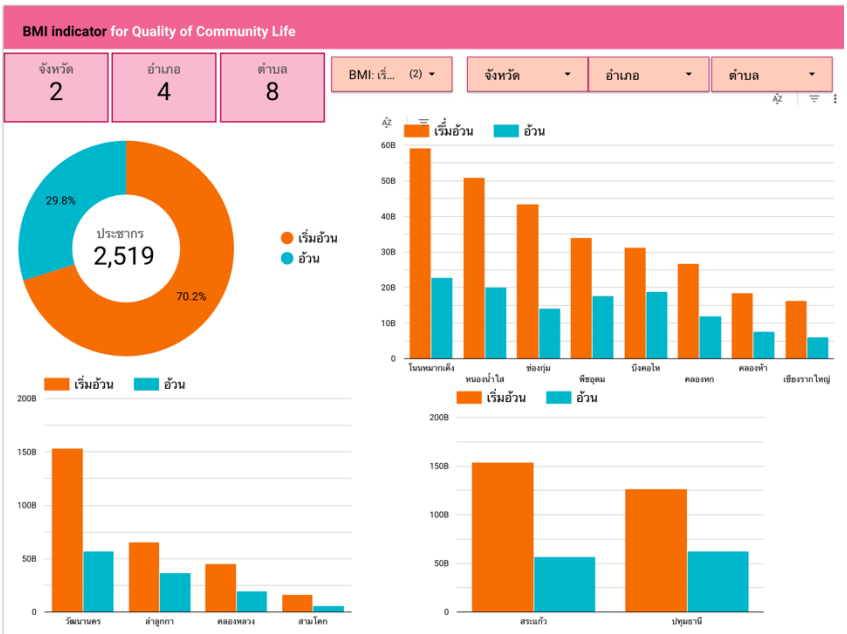
ค่าดัชนีมวลกาย BMI เป็นปัจจัยที่ใช้เป็นตัวชี้วัดทางด้านคุณภาพชีวิต (Simmonds et al., 2016) ได้ทำการศึกษา BMI ในการแสดงให้เห็นการก่อให้เกิดสภาวะอ้วนในวัยเด็กที่มีจำนวนเพิ่มมากขึ้น และส่งผลต่อการเกิดสภาวะอ้วนในอนาคต ดังนั้นงานวิจัยนี้จากภาพที่ 3 ได้ให้นำข้อมูลมาวิเคราะห์ เป็นตัวอย่าง กลุ่มที่เป็นความเสี่ยงในการเข้าสู่สภาวะอ้วนในแต่ละพื้นที่ มีจำนวนสูงถึง 2519 คน จากประชากร 8756 คนคิดเป็นร้อยละ 28.76 จะเห็นได้ว่า ลักษณะประชากรที่ถูกแบ่งออกเป็น สถานะปกติ เริ่มอ้วน ผอมเกินไป และอ้วนของแต่ละชุมชนในท้องถิ่น มีความแตกต่างกัน ตามสภาพของครัวเรือน รายได้ และเศรษฐกิจที่แตกต่างกันของแต่ละชุมชน ซึ่งในการทดลองนี้เราสามารถที่จะ ประมาณการของกลุ่มของพายุต่าง ๆ โดยเฉพาะอย่างยิ่งที่อยู่ในช่วงของเด็ก หรือวัยรุ่น ที่จะกลายเป็นผู้ใหญ่ในอนาคต ด้วยการคำนวณหาค่าดัชนีมวลกาย และคาดการณ์ จำนวนของผู้ที่เข้าสู่สภาวะเริ่มอ้วนหรืออ้วน ของผู้ใหญ่ในอนาคต ได้อย่างถูกต้องและแม่นยำ

ตัวอย่างการนำค่า BMI ไปใช้ในการตรวจวัดสุขภาพ โดยมุ่งเน้นการแก้ไขปัญหาโรคอ้วนในเด็กโดยใช้วิธีการทำงานร่วมกันในชุมชน โปรแกรม GO (Lek et al., 2021) ได้รับการออกแบบมาเพื่อช่วยเหลือเด็กที่มีน้ำหนักเกินหรือเป็นโรคอ้วนจากพื้นที่ที่มีความหลากหลายทางชาติพันธุ์และสถานะทางเศรษฐกิจต่ำ ด้วยการติดตามเด็กและวัยรุ่นที่มีน้ำหนักเกินหรือเป็นโรคอ้วนอายุ 4-19 ปี เป็นระยะเวลาสองปี**จาก**เด็ก 178 คนจากเขต Malburgen ในเมือง Arnhem ประเทศเนเธอร์แลนด์

นอกจากนี้ โปรแกรม GO! ซึ่งมีโค้ชสุขภาพเด็กเป็นศูนย์กลางในการประสานงานและให้คำแนะนำเกี่ยวกับการใช้ชีวิตอย่างมีสุขภาพ ผลที่ได้เปลี่ยนแปลงในคะแนน BMI-SDS (ดัชนีมวลกายมาตรฐาน) และคะแนน HRQoL (คุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ) โดยค่า BMI-SDS หลังจาก 24 เดือน ผู้ที่เข้าร่วมโปรแกรมอย่างต่อเนื่องมี BMI-SDS ลดลงมากกว่าผู้ที่เข้าร่วมไม่ต่อเนื่อง ดูจากระดับความอ้วน มีแนวโน้มไปในทิศทางที่ดีขึ้น โดยมีเด็กบางส่วนลดระดับความอ้วนลง สำหรับค่า HRQoL มีแนวโน้มปรับปรุงดีขึ้นโดยไม่คำนึงถึงอัตราการเข้าร่วม เพศ และเชื้อชาติ และพบว่า โปรแกรม GO! อาจมีประสิทธิภาพในการลด BMI-SDS ในเด็กที่มีน้ำหนักเกินหรือเป็นโรคอ้วนจากพื้นที่ที่มีความหลากหลายทางชาติพันธุ์และสถานะทางเศรษฐกิจต่ำ นอกจากนี้ยังมีแนวโน้มที่จะช่วยปรับปรุงคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพของเด็กอีกด้วย อย่างไรก็ตาม จำเป็นต้องมีการศึกษาเพิ่มเติมเพื่อยืนยันผลการวิจัยเหล่านี้

นอกจากนี้ การประกอบอาชีพ ระดับการศึกษาของสมาชิกครอบครัว จะส่งมีผลกระทบต่อ ค่าของดัชนีมวลกาย เนื่องจากว่ากระบวนการการศึกษานี้ มีเป้าหมายในการที่เราสามารถจะนำองค์ความรู้ต่าง ๆ ที่เรามีรูปแบบของการคิดวิเคราะห์ หรือการสื่อสารต่าง ๆ ที่เกิดจากการเรียนรู้ทั้งในระบบและนอกระบบ มาใช้ในการช่วยสนับสนุน และแก้ไขปัญหาที่เกิดขึ้น ทั้งระยะสั้นและระยะยาวได้

อีกตัวอย่างงานวิจัย (Myers et al., 2021) ที่ศึกษาความสัมพันธ์ระหว่างลักษณะงานกับดัชนีมวลกาย (BMI) และความเสี่ยงต่อโรคอ้วน โดยใช้ข้อมูลจากแบบสำรวจคุณภาพชีวิตการทำงานของสหรัฐอเมริกาปี 2014 พบว่า การทำงานกะกลางคืนและการทำงานประเภทแรงงานมีความสัมพันธ์เชิงบวกกับ BMI ที่สูงขึ้นและความเสี่ยงต่อโรคอ้วน ในขณะที่ลักษณะงานอื่น ๆ ไม่พบความสัมพันธ์ที่มีนัยสำคัญทางสถิติ จากผลดังกล่าวสรุปได้ว่า การทำงานกะกลางคืนและการทำงานประเภทแรงงานเป็นปัจจัยเสี่ยงต่อภาวะ BMI สูงและโรคอ้วน การระบุปัจจัยเสี่ยงในงานประเภทแรงงาน และการออกแบบงานใหม่เพื่อลดปัจจัยเสี่ยงเหล่านั้น รวมถึงการลดการทำงานกะกลางคืน อาจมีบทบาทในการลดความชุกของโรคอ้วนในสหรัฐอเมริกา



ภาพที่ 3 แสดง BMI ภาวะความเสี่ยงด้านสุขภาพ เริ่มอ้วน และ อ้วน ของพื้นที่ต่าง ๆ

ในการวิเคราะห์สุขภาพจากค่า BMI ในงานวิจัยชิ้นนี้จะเน้นสถานะที่เริ่มจากเข้าสู่ภาวะความเสี่ยง เนื่องจากเป็นปัจจัยที่ทำให้ภาพรวมคุณภาพของพื้นที่ มหาวิทยาลัยและชุมชนจะต้องหาแนวทางในการป้องกัน จากตัวอย่างข้างต้นเราทำการกรองข้อมูลเฉพาะในส่วนของ ค่าของข้อมูลที่มีความเสี่ยง เป็นส่วนสำคัญที่ใช้ในการให้ความสำคัญ ถือเป็นปัจจัยที่นำมาทดสอบ ก็คือ เริ่มอ้วน อ้วน ชุมชนในท้องถิ่น ในจังหวัดสระแก้วและปทุมธานีมีจำนวนผู้ที่เริ่มเข้าสู่ สภาวะอ้วน ถึง 70.2% ในขณะเดียวคนที่ อยู่ในสถานะอ้วนอยู่แล้วอยู่ที่ 29.8% การดูแลและให้ความสำคัญในการ ส่งเสริมด้วยกิจกรรมเพื่อพัฒนาสุขภาพโดยเร็ว เนื่องจากมีจำนวนของ ผู้ที่อยู่ในสภาวะ อ้วน สูงสุด ก็คืออำเภอวัฒนานคร จังหวัดสระแก้ว และ อำเภอลำลูกกา จังหวัดปทุมธานี ตามลำดับ

การเกิดภาวะเสี่ยง โรคภัยไข้เจ็บต่าง ๆ หรือ การเจ็บป่วย การได้รับผลข้อมูลการเกิดความ เสี่ยงต่าง ๆ ล่วงหน้า ถือว่าเป็นข้อดีที่ทำมหาวิทยาลัย สามารถที่จะจัดโครงการหรือกิจกรรม ที่รองรับและดูแลรักษา ปรับเปลี่ยนลดจำนวนผู้ที่มีสภาวะเริ่มอ้วน เพื่อจะได้ไม่เป็นสภาวะอ้วนในอนาคต เพิ่มคุณภาพชีวิตให้กับชุมชนได้เป็นอย่างดี ในลำดับต่อไปจะแสดงให้เห็นถึงขั้นตอนการเก็บรวบรวมข้อมูลเพื่อใช้สำหรับระบบปัญญาประดิษฐ์ ลดความเสี่ยงและเพิ่มความเข้มแข็งทางด้านสุขภาพให้กับชุมชนในท้องถิ่น

3. การเก็บรวบรวมข้อมูล

หลังจากที่มีการออกแบบลักษณะของข้อมูลประเภทของข้อมูล (Dataset Design) ที่ทำหน้าที่ในการจัดเก็บในรูปแบบต่าง ๆ และมีการทดสอบถึงข้อมูลที่จะนำไปใช้ในการวิเคราะห์ด้วยระบบปัญญาประดิษฐ์ ขั้นตอนต่อไปก็คือการลงพื้นที่ในการจัดเก็บข้อมูล เพื่อทำการตรวจสอบลักษณะของข้อมูลที่ใช้งานว่าตรงตามความต้องการที่กำหนด หรือไม่และนำไปใช้ในการวิเคราะห์ข้อมูล ได้ถูกต้องหรือไม่ ที่สำคัญยังกำหนดถึงวิธีการตรวจสอบความถูกต้องของข้อมูลในแต่ละขั้นตอน ซึ่งถือว่าเป็นกระบวนการที่สำคัญ หากข้อมูลที่ได้รับไม่ถูกต้องหรือไม่ รวมไปถึงปรับปรุงแก้ไข การประมวลผลและการวิเคราะห์ ในขั้นตอนต่อไป ก็จะทำได้ยากยิ่งขึ้น

ในงานวิจัยชิ้นนี้ เพื่อได้รับข้อมูลที่ถูกต้องมากที่สุด ดังนั้น บุคลากรที่ทำหน้าที่ในการใกล้ชิดกับท้องถิ่นมากที่สุดก็คืออาสาสมัครดูแลสุขภาพ (อ.ส.ม.) ในโรงพยาบาลส่งเสริมสุขภาพตำบล ช่วงเวลาในการลงพื้นที่จัดเก็บข้อมูลจะแบ่งออกเป็น 2 ช่วงด้วยกันคือขั้นตอนของการเข้าไปทำความเข้าใจ (Understanding phase) อธิบายรูปแบบ และประเภทของข้อมูลที่ได้รับการออกแบบ ก็จะทำการจัดเก็บข้อมูลในรอบที่หนึ่ง โดยให้ระยะเวลาในการจัดเก็บข้อมูล จำนวน 3-4 เดือน (พฤศจิกายน - กุมภาพันธ์) เมื่อได้รับข้อมูลเสร็จเรียบร้อย ทางมหาวิทยาลัยราชภัฏจะนำข้อมูลไปตรวจสอบ ความถูกต้องของข้อมูลและบันทึกลงในระบบฐานข้อมูล โดยจะใช้ระยะเวลาทั้งสิ้น จำนวน 1 เดือน (มีนาคม) จากการทดลองพบว่า ช่วงเวลาของการจัดเก็บข้อมูลถือว่าเป็นเรื่องสำคัญ จำเป็นจะต้องทิ้งระยะเวลาในการจัดเก็บข้อมูลหรือการพักการจัดเก็บข้อมูล เพื่อให้ข้อมูลมีความถูกต้องและผู้ให้ข้อมูลสามารถที่จะเข้าใจและเรียนรู้ถึงข้อมูลที่จัดเก็บ ในรอบที่ผ่านมา อย่างไรก็ตามมีข้อควรระวังก็คือหากทิ้งระยะเวลาในการเก็บข้อมูลจำนวนมากเกินไป การขาดความต่อเนื่องจะสูญเสียข้อมูลในช่วงเดือนที่ไม่ได้ลงไปทำการจัดเก็บข้อมูล อาทิ ข้อมูลที่ใช้ในการติดตามค่าดัชนีมวลกายต่าง ๆ มีผลต่อการเปลี่ยนแปลงเพิ่มขึ้นหรือลดลงในแต่ละปี ไม่ว่าจะเป็นน้ำหนักส่วนสูงหรืออายุ เป็นตัวอย่างที่แสดงถึง ความเป็นอยู่ของแต่ละชุมชน ซึ่งจะมีผลต่อการคำนวณของค่า BMI

ขั้นตอนการลงพื้นที่ ในการจัดเก็บข้อมูลในรอบที่สอง ซึ่งระยะเวลาในขั้นตอนที่สองนี้จะใช้ระยะเวลา จำนวน 3 ถึง 4 เดือน (เมษายน - กรกฎาคม) เพื่อทบทวนความถูกต้องของข้อมูลและติดตามความก้าวหน้าของข้อมูลที่เกิดขึ้นในแต่ละปี ซึ่งขั้นตอนที่ยากที่สุดก็คือการบันทึกข้อมูลลงในฐานข้อมูล จำเป็น

จะต้องมีระยะเวลาในการทำงานที่เพิ่มมากขึ้น จำนวน 2 เดือน (สิงหาคม - กันยายน) จะรวมไปถึงเรื่องของการวิเคราะห์ การติดตามผลกระทบต่าง ๆ ที่เกิดขึ้น ในข้อมูลในแต่ละประเภท และในแต่ละมิติ

สำหรับการทดลองโดยงานวิจัยชิ้นนี้จะใช้โปรแกรม Google Data Studio ซึ่งปัจจุบัน ได้เปลี่ยนเป็นโปรแกรมชื่อ Looker Studio ที่พัฒนาโดยกูเกิล Google Application ยอมรับการประมวลผลข้อมูล และนำเสนอข้อมูล (Data Visualization) มีความสามารถในการเชื่อมต่อกับโปรแกรมฐานข้อมูล หรือ สเปรดชีต Spreadsheet อย่างสะดวกและรวดเร็ว ในงานวิจัยนี้ใช้โปรแกรม Orange กำหนดสูตรสมการในการคำนวณต่าง ๆ ลงไปในช่วงข้อมูลที่ต้องการ กระบวนการในการวิเคราะห์ต่าง ๆ จะต้องหา ความแม่นยำ ในการคำนวณและคาดการณ์จำนวนของผู้ที่อยู่ในสถานะต่าง ๆ ของชุมชน ตามค่าดัชนีมวลกายที่เกิดขึ้นได้อย่างแม่นยำ โดยในงานวิจัยชิ้นนี้ใช้สมการ (Machine Learning algorithms) หลายรูปแบบเพื่อประกอบการคำนวณและเลือกใช้สมการที่มีค่าความแม่นยำมากที่สุดในการทำงาน

4. การวิเคราะห์ข้อมูล

ในขั้นตอนการวิเคราะห์ งานวิจัยนี้นำระบบปัญญาประดิษฐ์มาสนับสนุนการทำนายผลค่าดัชนีมวลกาย (BMI) ของชุมชนตามหลักการของการเรียนรู้ของเครื่อง (Machine Learning) ค่า BMI ที่ได้เป็นปัจจัยที่ใช้ในการจำแนกข้อมูล (Classification) โดยอาศัยจากข้อมูลท้องถิ่นของกลุ่มข้อมูลราชภัฏที่มีการรวบรวมจากตำบลต่าง ๆ ข้อมูลเปรียบเทียบข้อมูลท้องถิ่นรวบรวมมาก่อนหน้านั้น การวิเคราะห์จะแบ่งกลุ่มข้อมูลออกเป็น 2 ส่วนคือข้อมูลสำหรับการเรียนรู้ (Learning data) และ กลุ่มข้อมูลสำหรับการทดสอบ(Testing data) เพื่อมาเพิ่มประสิทธิภาพในการออกแบบข้อมูลที่เกี่ยวข้องให้ถูกต้องและผลลัพธ์ที่แม่นยำ ขึ้นกับข้อมูลที่ครบถ้วนและสมบูรณ์ สนับสนุนในการบริหารทิศทางโครงการต่าง ๆ ให้สอดคล้องกับคุณภาพชีวิตของท้องถิ่นและชุมชน

เริ่มการวิเคราะห์ข้อมูล ค้นหาปัจจัยที่ใช้ในการคำนวณค่าดัชนีมวลกายจาก ส่วนสูง (ซม.) และน้ำหนัก (กก.) โดยการติดตามคุณภาพชีวิตได้จาก ช่วงอายุของประชากรและเพศที่ก่อให้เกิดโอกาสของผลลัพธ์ในลักษณะต่าง ๆ อาทิ โอกาสของคนในชุมชนที่จะมี “เริ่มอ้วน” หรือ ”คนอ้วน” เพิ่มมากขึ้น เป็นต้น เรียกกันว่า Model ที่ใช้ในการวิเคราะห์ตามคุณลักษณะ (Feature) และคำนวณหา ผลลัพธ์ Target หรือ Class ที่ต้องการได้

	Name	Type	Role	Values
1	เพศ	C categorical	feature	ชาย,หญิง
2	อายุ	N numeric	feature	
3	ส่วนสูง(ซม.)	N numeric	feature	
4	น้ำหนัก(กก.)	N numeric	feature	
5	Result	C categori...	target	ปกติ,ผอมเกินไป,อ้วน,เริ่มอ้วน
6	BMI	S text	meta	
7	รหัสแบบสอบถ...	S text	meta	

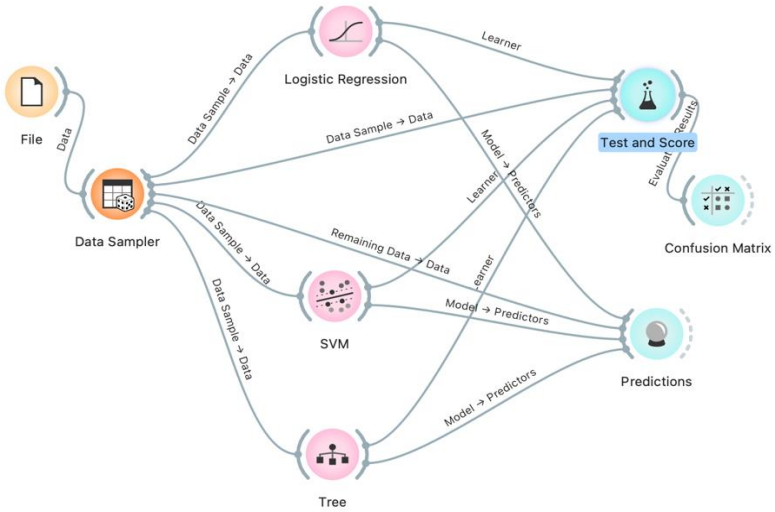
ภาพที่ 4 แสดงการกำหนดปัจจัยที่สำคัญ (Feature) และผลลัพธ์ (Class หรือ Target)

จากภาพที่ 4 เป็นวิธีการจำแนกข้อมูลเพื่อวิเคราะห์ข้อมูลที่สำคัญ ในการวิจัยชิ้นนี้ได้กำหนด ปัจจัยที่มีผลต่อการ เกิดความเสี่ยง ได้ทำการวิเคราะห์ จาก เพศ อายุ ส่วนสูงและน้ำหนัก เราจะใช้เป็น “Feature” เพื่อใช้ในการประมวลผล ผ่านการเรียนรู้ของเครื่อง และทดสอบ โดยทำการแบ่งข้อมูลให้เครื่องเรียนรู้จำนวน

80 เปอร์เซ็นต์ เรียกว่า ข้อมูลเพื่อการเรียนรู้ “Training Data” และที่เหลืออีก 20 เปอร์เซ็นต์ จะใช้เป็นข้อมูลทดสอบความถูกต้อง หรือ “Testing Data”

ระบบปัญญาประดิษฐ์จะคาดคะเนผลลัพธ์ Results ว่า ประชากรชุมชนมีความเสี่ยง หรือโอกาสที่จะเกิดโรคภัยไข้เจ็บ ในระดับใด โดยแบ่งเป็น อ้วน เริ่มอ้วน ปกติ หรือผอมเกินไป โดยกำหนดเป็น Category Target หลังจากนั้นก็เข้าสู่ขั้นตอนในการคัดเลือกโมเดลที่สำคัญเพื่อที่จะหาคนในการทำนายที่ใกล้เคียงความเป็นจริงมากที่สุด ซึ่งรูปแบบนี้เราได้ เลือกใช้ Model จากสมการจำนวน 3 รูปแบบด้วยกัน คือ

- Logistic Regression Model
- SVM Model
- Decision Tree Model



ภาพที่ 5 วิเคราะห์ด้วย Machine Learning: Logistic Regression SVM และ Tree

จากภาพที่ 5 เป็นการดำเนินการขั้นการนำข้อมูลเข้าไปทำงานตามสมการต่าง ๆ มีการกำหนด ข้อมูลนำเข้า (Data Sample) ต่อไปก็คือการค้นหารูปแบบของการวิเคราะห์ข้อมูลในวิธีต่าง ๆ Logistic Regression SVM และ Tree ตามแผนภาพข้างต้น ผลลัพธ์ที่ได้จะจัดเก็บที่ ตารางข้อมูล(Test and Score) ในรูปแบบตาราง เพื่อเปรียบเทียบค่าของข้อมูลต่าง ๆ Confusion Matrix รูปแบบของวิธีใดที่มีความเหมาะสมที่สุดที่จะใช้ในการเป็นตัววิเคราะห์และประมวลผลข้อมูลต่าง ๆ โดยที่ การทำนายผลที่มีความชัดเจนจะมีค่าเข้าใกล้ 1 มากที่สุด

ตารางที่ 1 Confusion Matrix

การทำนาย (Prediction)	**Actually Positive (มีค่า = 1)**	**Actually Negative (มีค่า = 0)**
Predicted Positive (มีค่า = 1)	True Positives (TPs)	False Positives(FPs)
Predicted Negative(มีค่า = 0)	False Negatives (FNs)	True Negatives(TNs)

ในการทำนาย ค่าความถูกต้อง (Accuracy) เกิดจากผลรวมของ TPs และ TNs หารด้วย ผลรวมของค่าที่ได้ทั้งหมด (TPs+FPs+FNs+TNs) และคำนวณค่าความแม่นยำ (Precision) ได้จากการการทำนายว่าเกิดขึ้นจริง และผลที่ได้มีการเกิดขึ้นจริงจาก TPs/(TPs+FPs) ท้ายที่สุด การคำนวณหาความถูกต้องของการทำนายว่าเกิดขึ้นจริง (Recall) จาก TPs/(TPs+FNs) ในงานวิจัยนี้ จะใช้รูปแบบตารางนี้ เพื่ออธิบายความแม่นยำของสมการต่าง ๆ โดยข้อมูลประชากรที่อยู่ในชุมชนดูจากค่าดัชนีมวลกาย(BMI) ที่ได้จากส่วนสูงและน้ำหนัก โดยมีการติดตามช่วงอายุของประชากร เพื่อหาโอกาสของจำนวนประชากรตามอายุต่าง ๆ ที่เกิดขึ้น ลำดับต่อไปจะอธิบายหลักวิธีการของสมการที่ใช้ในการงานวิจัยนี้

หลักการวิเคราะห์ข้อมูลในรูปแบบ Logistic Regression Model

สำหรับหลักการวิเคราะห์การถดถอยแบบโลจิสติกส์ ซึ่งเป็นเทคนิควิธีการ ในการใช้ในการทำนาย สิ่งที่จะเกิดขึ้นกับสิ่งที่ศึกษาว่ามีความเป็นไปได้ ผลลัพธ์จะอยู่ในรูปแบบ 2 คำตอบหรือ 2 ทิศทาง (Binary outcome) การทำนายดังกล่าวจะอาศัยข้อมูล หรือปัจจัยอื่น ๆ มาประกอบการพิจารณา อาทิเช่น หากเราต้องการจะทราบว่านักศึกษาจะสอบผ่านหรือไม่ เราต้องอาศัยข้อมูลผลการเรียน คะแนนสอบ และปัจจัยต่าง ๆ มาประกอบเพื่อทำนายผลการสอบผ่านหรือไม่ผ่านของนักศึกษาแต่ละคน นอกจากนี้ ยังใช้วิธีการนี้ในการประเมินว่าจะเกิดสภาวะหัวใจวายของผู้ป่วยหรือไม่ โดยดูจากอายุ เพศ ประวัติของครอบครัว ระดับคอเลสเตอรอลและปัจจัยแวดล้อมมาประเมินด้วย

สำหรับในงานวิจัยนี้ ข้อมูลท้องถิ่นที่รวบรวมจากพื้นที่มาคำนวณหาและวิเคราะห์หาดัชนีมวลกาย (Body Mass Index) โดยกำหนดเบื้องต้นก็คือเป็นข้อมูลที่เกิดจากปัจจัยตัวแปร ในการคำนวณ ซึ่งผลในพื้นที่ต่าง ๆ ทำให้เราทราบถึงข้อมูลที่เกิดขึ้นว่า ทำนายเหตุการณ์หรือประเมินความเสี่ยงที่จะเกิดขึ้น มีความเป็นไปได้ที่จะเสี่ยงต่อ คุณภาพชีวิตของคนในชุมชนว่า มีโอกาส หรือความเสี่ยงความอ้วน มากน้อย หรือแม้กระทั่ง ไม่อ้วนและปกติ เป็นต้น ด้วยการใช้วิธีการประมาณความน่าจะเป็นสูงสุด (Maximum Likelihood Estimation)

นอกจากนี้ในตัวอย่างการศึกษา (Simmonds et al., 2016) ที่ชี้ให้เห็นว่าภาวะอ้วนในวัยเด็กเป็นปัจจัยเสี่ยงสำคัญต่อภาวะอ้วนในวัยผู้ใหญ่ ดังนั้น การดำเนินการเพื่อลดและป้องกันโรคอ้วนในวัยเด็กและวัยรุ่นจึงเป็นสิ่งจำเป็น อย่างไรก็ตาม การมุ่งเน้นการลดโรคอ้วนเฉพาะในเด็กที่เป็นโรคอ้วนหรือมีน้ำหนักเกินอาจไม่เพียงพอที่จะลดภาระโดยรวมของโรคอ้วนในผู้ใหญ่ เนื่องจากผู้ใหญ่อ้วนส่วนใหญ่ไม่ได้เป็นโรคอ้วนในวัยเด็ก โดยมีผลการวิจัยเด็กและวัยรุ่นที่เป็นโรคอ้วน มีแนวโน้มที่จะเป็นโรคอ้วนในวัยผู้ใหญ่สูงกว่าผู้ที่ไม่อ้วนประมาณ 5 เท่า ประมาณ 55% ของเด็กอ้วนจะเป็นโรคอ้วนในวัยรุ่น ประมาณ 80% ของวัยรุ่นอ้วนจะยังคงเป็นโรคอ้วนในวัยผู้ใหญ่ และประมาณ 70% จะเป็นโรคอ้วนเมื่ออายุมากกว่า 30 ปี อย่างไรก็ตาม 70% ของผู้ใหญ่อ้วนไม่ได้เป็นโรคอ้วนในวัยเด็กหรือวัยรุ่น

หลักการวิเคราะห์ข้อมูล Support Vector Machine (SVM)

หลักการในการ จำแนกข้อมูล (Classification) รูปแบบ SVM เป็นที่ได้รับความนิยมเป็นอย่างยิ่งสำหรับวิธีการในการจำแนกข้อมูลของเครื่อง (Machine Learning Algorithm) ข้อมูลสำหรับการวิเคราะห์ลักษณะนี้ปริมาณข้อมูลในกลุ่มข้อมูลที่ไม่มาก ข้อมูลมีการกำหนดข้อมูลไว้เรียบร้อย ทำให้เหมาะในการจำแนกข้อมูล วิธีการก็คือ แบ่งข้อมูลออกเป็นกลุ่มด้วยการกำหนดเส้นแบ่งเป็น 2 เส้น แยกข้อมูลออกจากกัน เราเรียกเส้นนี้ว่าแนวแบ่งข้อมูล(Margin Classification) นี้มีความแคบมาก (Hard Margin Classification) เพื่อให้เกิดความชัดเจนและละเอียดถึงคุณลักษณะที่เกิดขึ้นข้อมูล บางข้อมูลที่ล้ำหรือปรากฏเข้ามาในแนวแบ่งดังกล่าว เราเรียกว่าเป็น

ความกว้างมาก (Soft Margin Classification) ของกลุ่มข้อมูลราชภัฏ ที่เกิดจากข้อมูลท้องถิ่นถูกจำแนกสถานะดัชนีมวลกายออกตามประเภทต่าง ๆ สามารถที่จะดำเนินการด้วยวิธีการ ตามหลักการของ SVM ได้เป็นอย่างดี

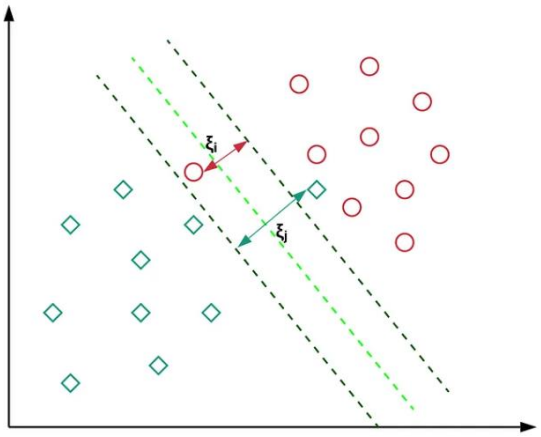
ตัวอย่างขั้นตอนการนำ SVM ไปใช้ในการวิเคราะห์ข้อมูล อาทิเช่น ในงานวิจัย (Nanglia et al., 2021) ได้เสนออัลกอริทึมแบบผสมผสานที่เรียกว่า Kernel Attribute Selected Classifier (KASC) สำหรับการจำแนกประเภทมะเร็งปอด โดยใช้วิธีการ Support Vector Machine (SVM) และ Feed-Forward Back Propagation Neural Network (FFBPNN) ร่วมกัน โดยที่KASC ถูกนำไปใช้กับชุดข้อมูลภาพปอด ELCAP ซึ่งประกอบด้วยภาพ CT scan ความละเอียดต่ำ 500 ภาพ กระบวนการจำแนกภาพประกอบด้วย 3 ขั้นตอนหลัก ได้แก่

1. การประมวลผลข้อมูลล่วงหน้า (Data preprocessing) ภาพจะถูกแบ่งส่วนโดยใช้การดำเนินการทางสัณฐานวิทยาและตรรกะคลุมเครือ จากนั้นจึงทำการสกัด Region of Interest (ROI)

2. การสกัดและปรับแต่งคุณลักษณะ (Feature extraction and optimization) มีการใช้เทคนิค Speeded Up Robust Features (SURF) เพื่อสกัดคุณลักษณะจาก ROI และใช้อัลกอริทึมทางพันธุกรรม (Genetic Algorithm - GA) เพื่อปรับแต่งคุณลักษณะที่สกัดได้

3. การผสมผสาน SVM และ NN เพื่อการทำนาย (Hybridization of SVM and NN for prediction) โดย SVM ที่มีเคอร์เนลแบบ polynomial จะถูกใช้เพื่อเลือกคุณลักษณะที่เหมาะสมที่สุด จากนั้นคุณลักษณะที่เลือกนี้จะถูกส่งต่อไปยัง FFBPNN เพื่อทำการจำแนกประเภท

ผลการทดลองแสดงให้เห็นว่า KASC ที่ใช้เคอร์เนลแบบ polynomial มีความแม่นยำในการจำแนกประเภทสูงกว่า KASC ที่ใช้เคอร์เนลแบบ linear และสูงกว่างานวิจัยก่อนหน้า โดยมีความแม่นยำเฉลี่ยอยู่ที่ 98.08% เมื่อใช้ตัวอย่างฝึก 500 ตัวอย่าง และสรุปได้ว่า KASC เป็นวิธีการที่มีประสิทธิภาพสำหรับการจำแนกมะเร็งปอด ซึ่งเกิดจากการรวมกันของคุณสมบัติการสกัดคุณลักษณะที่มีประสิทธิภาพของ SURF การปรับแต่งคุณลักษณะที่มีประสิทธิภาพของ GA และการผสมผสานระหว่าง SVM ที่ใช้เคอร์เนลแบบ polynomial กับ FFBPNN



ภาพที่ 6 แนวทางการปรับเส้นแบ่งกลุ่มข้อมูล Hard/Soft Margin SVM

รูปแบบ SVM นี้ มีการใช้งานกันอย่างแพร่หลาย โดยเฉพาะอย่างยิ่งการนำเอาวิธีการไปใช้ในการประยุกต์ เรื่องของการดูแลผู้ป่วยมะเร็งทรวงอก ผลการวินิจฉัยโรคมะเร็งมีความแม่นยำถึง 94% หรือแม้กระทั่งจะมีการนำเอาวิธีการนี้ไปใช้ในการคำนวณเพื่อหาความเสี่ยงที่เกิดโรคภาวะของการเกิด Stroke นอกจากนี้ยัง (Nanglia et al., 2021) นำเสนอศึกษาใช้วิธีการ SVM ในการคัดกรองผู้ป่วย ผู้ป่วยโรคมะเร็งในปอด ที่เหมาะต่อการให้เคมีบำบัดเพื่อได้รับประสิทธิภาพมากที่สุด

วิธีการนี้ยังนำไปใช้ในการทำการเกษตรกรรม ตัวอย่างงานวิจัย (Jena et al., 2021) เป็นตัวอย่างของการมุ่งเน้นการจำแนกโรคข้าวที่สำคัญ เช่น โรคใบจุดสีน้ำตาล โรคหนอนห่อใบข้าว และโรคไหม้ โดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ที่หลากหลาย เช่น k-NN, neural network, SVM, decision tree, random forest, Logistic Regression, AdaBoost และ Naïve Bayes เพื่อวิเคราะห์ภาพถ่ายของโรคข้าว และแยกแยะโรคต่าง ๆ รวมถึงกรณีต้นข้าวปกติ ผลการวิจัยชี้ให้เห็นว่าเทคนิคการเรียนรู้ของเครื่องมีประสิทธิภาพในการจำแนกโรคข้าว ซึ่งเป็นประโยชน์ต่อการวางแผนป้องกันและจัดการโรคข้าวได้อย่างทันท่วงที

สำหรับในงานวิจัยชิ้นนี้ นำวิธีการ SVM มาใช้งานมีความเหมาะสมในการแยกระหว่างจำนวนของผู้ที่เกิดสภาวะผอม เริ่มอ้วน และเริ่มอ้วน ในแต่ละระยะของช่วงอายุเพื่อติดตามคุณภาพชีวิตของ ประชากรที่อยู่ในชุมชนในท้องถิ่น นอกจากนี้สนับสนุนการจัดกิจกรรมต่าง ๆ เข้าไปสนับสนุนตามช่วงของกลุ่มที่เกิดขึ้น

หลักการวิเคราะห์ข้อมูลด้วยต้นไม้การตัดสินใจ (Tree)

ในการพัฒนาคุณภาพชีวิตของคนในชุมชน การวิเคราะห์ข้อมูลด้วยต้นไม้ตัดสินใจ ตัวอย่างงานวิจัย (Broda et al., 2021) ในงานวิจัย แสดงให้เห็นศักยภาพของวิธีการเรียนรู้ของเครื่อง (machine learning) ในการวิเคราะห์ข้อมูลเพื่อสนับสนุนการตัดสินใจเชิงนโยบายและแนวทางปฏิบัติที่ส่งผลกระทบต่อผู้ที่มีความบกพร่องทางสติปัญญาและพัฒนาการ (IDD) โดยใช้ข้อมูลจากแบบสำรวจ National Core Indicators In-Person Survey (NCI-IPS) ซึ่งเป็นแบบสำรวจประจำปีที่ได้รับการยอมรับในระดับประเทศของผู้ที่มี IDD มากกว่า 20,000 คน ทีมวิจัยได้สร้างแบบจำลองการจำแนกประเภทด้วยวิธี classification tree และ random forest เพื่อทำนายสถานะการจ้างงานและการเข้าร่วมกิจกรรมในแต่ละวันของบุคคล โดยใช้ข้อมูลจากคำตอบของพวกเขาในแบบสำรวจ NCI-IPS ปี 2017–2018 แบบจำลองที่แม่นยำที่สุดคือ random forest classifier ซึ่งสามารถทำนายผลลัพธ์การจ้างงานของผู้ใหญ่ที่มี IDD ได้อย่างแม่นยำถึง 89 เปอร์เซ็นต์ในกลุ่มตัวอย่างที่ใช้ทดสอบ และ 80 เปอร์เซ็นต์ในกลุ่มตัวอย่างที่กันไว้ ตัวแปรที่สำคัญที่สุดในการทำนายนี้คือว่าการจ้างงานในชุมชนเป็นเป้าหมายในแผนการบริการของบุคคลนี้หรือไม่ ผลลัพธ์เหล่านี้ชี้ให้เห็นถึงศักยภาพของเครื่องมือ machine learning ในการตรวจสอบผลลัพธ์ที่มีค่าอื่น ๆ ที่ใช้ในการตัดสินใจเชิงนโยบายบนหลักฐานเพื่อสนับสนุนผู้ที่มี IDD

การวิเคราะห์ข้อมูลด้วยต้นไม้การตัดสินใจ มีวิธีการ คือการจัดรูปแบบของข้อมูลการแบ่งข้อมูลตามกลุ่มข้อมูลต่าง ๆ โดยอาศัย สิ่งที่เรียกว่า การหาค่า (Value) ที่แบ่งเป็นลำดับ ยกตัวอย่างเช่นในหมู่บ้านแห่งหนึ่งมีข้อมูลการมีปัญหาทางสุขภาพ หากต้องการหาสาเหตุของการเกิดปัญหาสุขภาพ แนวทางที่ หาข้อมูลได้ เราข้อมูลจากสิ่งที่มีก็คือ การคำนวณในส่วนของการหาค่าดัชนีมวลกาย (BMI) ด้วยการแบ่งกลุ่มของข้อมูล ตามอายุ ตามน้ำหนัก เพื่อแบ่งกลุ่มข้อมูลท้องถิ่น อาทิ คนที่อ้วน อ้วนมาก อ้วนน้อย ไม่อ้วนหรือปกติ โอกาสที่จะเกิดปัญหาทางสุขภาพสามารถแยกแยะเลยจัดกลุ่มของข้อมูลได้อย่างชัดเจนว่า มีโอกาสที่จะเกิดปัญหาสุขภาพจากค่า BMI มีความ อ้วนเพิ่มมากขึ้น ของข้อมูล จะเป็นตัวบ่ง ถึงข้อมูลที่เกิดขึ้น

ผลการวิจัย

จากการทดลองจะพบว่าเมื่อนำเอารูปแบบของการเรียนรู้ของเครื่องมาใช้ในการประเมินผล เพื่อค้นหาโมเดลที่เหมาะสมสำหรับกันคาดการณ์หรือการทำนายผล อาจจะพบว่าโมเดลที่มี ค่าได้ใกล้เคียงสูงสุด โดยสามารถดูได้จากค่าของ CA หรือ “Classification Accuracy” ก็คือ Decision Tree ที่มีข่าวความแม่นยำ มาถึง 0.857 รองลงมาก็คือ Logistic Regression ที่มีค่า CA อยู่ที่ 0.834 และวิธี SVC มีค่า CA ที่ 0.832 ดังนั้นในคุณภาพชีวิต ด้าน BMI ลดความเสี่ยงของการเกิดโรคต่าง ๆ ของชุมชนในท้องถิ่น เกี่ยวกับเรื่องของการคำนวณ

ภาพรวมของประชากรหรือชุมชนที่อยู่ในท้องถิ่น จากการกำหนด เพศ ส่วนสูงและน้ำหนัก เลือกใช้โมเดล Decision Tree จะเหมาะสมที่สุด

Model	AUC	CA	F1	Precision	Recall
Tree	0.903	0.857	0.854	0.854	0.857
SVM	0.976	0.832	0.818	0.832	0.832
Logistic Regression	0.954	0.834	0.830	0.829	0.834

ภาพที่ 7 แสดงการทดสอบ ความถูกต้องแต่ละโมเดล (Test and Score)

อย่างไรก็ตาม ในการจัดหาข้อมูล เมื่อพิจารณา ในการเลือกโมเดล พบว่า โมเดล SVM กับ Decision Tree โดยจะเป็นคู่ที่ใช้ในการประมวลผล ที่มีค่าสูงสุดคือ 0.997 (ใกล้เคียง 1) ในขั้นตอนต่อไปเราจะทดลองเลือกใช้โมเดลทั้งสองตัวนี้ ในการประมวลผล เพื่อใช้ในการคาดคะเนผล ภาวะความเสี่ยงต่อโรคภัยไข้เจ็บของชุมชนในท้องถิ่นที่เกิดขึ้น

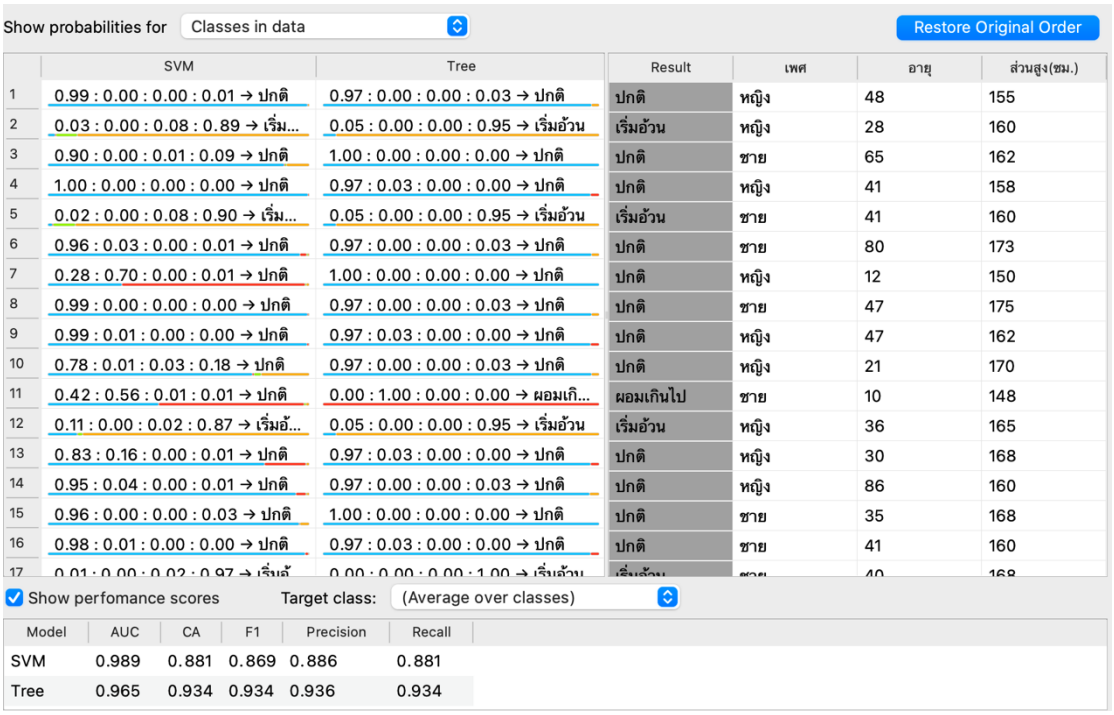
	Tree	SVM	Logistic Regression
Tree		0.003	0.018
SVM	0.997		0.946
Logistic Regression	0.982	0.054	

ภาพที่ 8 ผลของการวิเคราะห์ ค่าของโมเดล Decision Tree และ SVM มีค่าสูงสุด 0.997

ซึ่งผลจากการวิเคราะห์จะเห็นได้ว่าจากภาพที่ 7 ในการใช้โมเดลของการวิเคราะห์ SVM และ Tree จะมีค่าถึง 0.977 ซึ่งถือว่าเป็นวิธีการที่มีความเหมาะสม ทั้งสองรูปแบบนี้ในส่วนของ Logistic Regression และ Decision Tree ก็ยังคง มีค่า 0.982 ยังเป็นที่น้อยกว่า ดังนั้น การทำงานในการวิเคราะห์ข้อมูลในก็จะเลือกใช้ SVM และ Decision Tree

อภิปรายผล

จากการผลการทดลองในตัวอย่างการหารูปแบบของ การวิเคราะห์คุณภาพชีวิตตาม ค่า BMI ของข้อมูลท้องถิ่น ยกตัวอย่างเช่น ในภาพจะเห็นว่า ในบรรทัดที่ 11 การใช้โมเดล SVM ทำนายว่า “ปกติ” แต่ในส่วนของ Decision Tree Model ทำนายว่า”ผอมเกินไป” ซึ่งเป็นคำตอบที่ถูกต้อง



Show probabilities for: Classes in data — Restore Original Order

	SVM	Tree	Result	เพศ	อายุ	ส่วนสูง(ซม.)
1	0.99 : 0.00 : 0.00 : 0.01 → ปกติ	0.97 : 0.00 : 0.00 : 0.03 → ปกติ	ปกติ	หญิง	48	155
2	0.03 : 0.00 : 0.08 : 0.89 → เริ่ม...	0.05 : 0.00 : 0.00 : 0.95 → เริ่มอ้วน	เริ่มอ้วน	หญิง	28	160
3	0.90 : 0.00 : 0.01 : 0.09 → ปกติ	1.00 : 0.00 : 0.00 : 0.00 → ปกติ	ปกติ	ชาย	65	162
4	1.00 : 0.00 : 0.00 : 0.00 → ปกติ	0.97 : 0.03 : 0.00 : 0.00 → ปกติ	ปกติ	หญิง	41	158
5	0.02 : 0.00 : 0.08 : 0.90 → เริ่ม...	0.05 : 0.00 : 0.00 : 0.95 → เริ่มอ้วน	เริ่มอ้วน	ชาย	41	160
6	0.96 : 0.03 : 0.00 : 0.01 → ปกติ	0.97 : 0.00 : 0.00 : 0.03 → ปกติ	ปกติ	ชาย	80	173
7	0.28 : 0.70 : 0.00 : 0.01 → ปกติ	1.00 : 0.00 : 0.00 : 0.00 → ปกติ	ปกติ	หญิง	12	150
8	0.99 : 0.00 : 0.00 : 0.00 → ปกติ	0.97 : 0.00 : 0.00 : 0.03 → ปกติ	ปกติ	ชาย	47	175
9	0.99 : 0.01 : 0.00 : 0.00 → ปกติ	0.97 : 0.03 : 0.00 : 0.00 → ปกติ	ปกติ	หญิง	47	162
10	0.78 : 0.01 : 0.03 : 0.18 → ปกติ	0.97 : 0.00 : 0.00 : 0.03 → ปกติ	ปกติ	หญิง	21	170
11	0.42 : 0.56 : 0.01 : 0.01 → ปกติ	0.00 : 1.00 : 0.00 : 0.00 → ผอมเกิ...	ผอมเกินไป	ชาย	10	148
12	0.11 : 0.00 : 0.02 : 0.87 → เริ่มอ้...	0.05 : 0.00 : 0.00 : 0.95 → เริ่มอ้วน	เริ่มอ้วน	หญิง	36	165
13	0.83 : 0.16 : 0.00 : 0.01 → ปกติ	0.97 : 0.03 : 0.00 : 0.00 → ปกติ	ปกติ	หญิง	30	168
14	0.95 : 0.04 : 0.00 : 0.01 → ปกติ	0.97 : 0.00 : 0.00 : 0.03 → ปกติ	ปกติ	หญิง	86	160
15	0.96 : 0.00 : 0.00 : 0.03 → ปกติ	1.00 : 0.00 : 0.00 : 0.00 → ปกติ	ปกติ	ชาย	35	168
16	0.98 : 0.01 : 0.00 : 0.00 → ปกติ	0.97 : 0.03 : 0.00 : 0.00 → ปกติ	ปกติ	ชาย	41	160
17	0.01 : 0.00 : 0.02 : 0.97 → เริ่มอ้...	0.00 : 0.00 : 0.00 : 1.00 → เริ่มอ้วน	เริ่มอ้วน	ชาย	40	168

Show perfomance scores — Target class: (Average over classes)

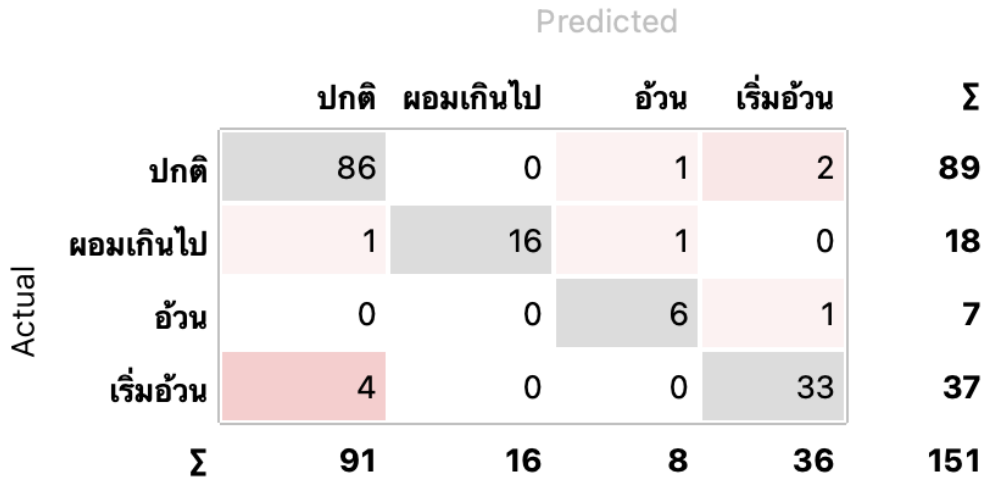
Model	AUC	CA	F1	Precision	Recall
SVM	0.989	0.881	0.869	0.886	0.881
Tree	0.965	0.934	0.934	0.936	0.934

ภาพที่ 9 ผลลัพธ์ของการทำนายด้วยโมเดล Logistic Regression SVM และ Decision Tree

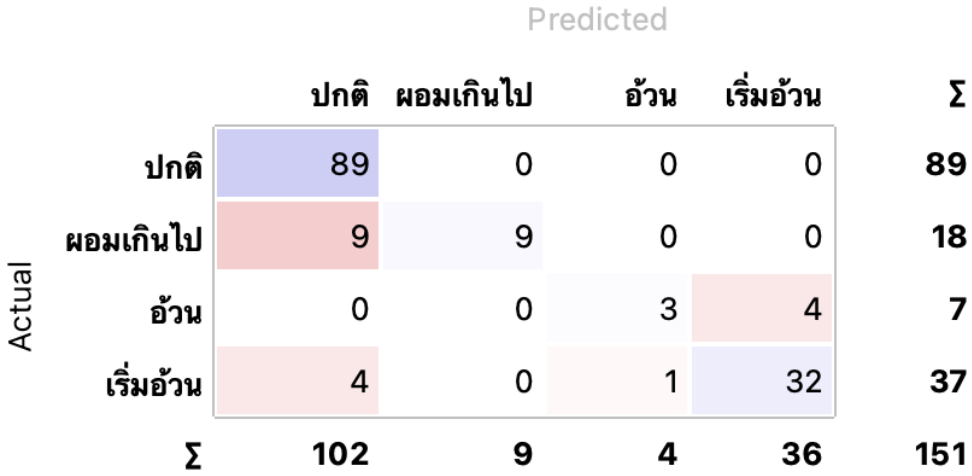
การเลือกใช้โมเดล จำเป็นจะต้องอาศัยทักษะ วิเคราะห์ เพื่อให้ได้ผลลัพธ์ที่สมบูรณ์มากที่สุด หรือเราเรียกกันว่า มีความผิดพลาด น้อยที่สุดให้เกิด จากงานวิจัยชิ้นนี้เราจะใช้ การดูจาก ตาราง Confusion Matrix ที่ทำการเปรียบเทียบวิธีการในแต่ละโมเดล เพื่อให้เห็นถึงผลลัพธ์ที่ถูกต้อง ตามแนวทแยงของตารางของโมเดล เมื่อนำข้อมูลมาทดสอบ ผลที่ได้จะเป็นไปตามตาราง ดังกล่าว

ตารางจากภาพที่ 9 จะเห็น ผลการทำนายจากโมเดล Decision Tree Model ว่าข้อมูลที่เกิดขึ้น เพื่ออธิบายความแม่นยำ จากตารางดังกล่าว แนวทแยงของตาราง แสดงความถูกต้องของข้อมูลที่เกิดขึ้น แนวตั้งแสดงถึงการทำนายจากโมเดล (Predicted) ส่วนแนวนอน เป็นค่าของข้อมูลหรือผลที่เกิดขึ้นจริง (Actual) อาทิ จากข้อมูล 89 คนที่ค่า BMI จริงระบุว่า “ปกติ” ในการทำนายทำนายว่า ปกติ ได้ถูกต้องจำนวน 86 และทำนายไม่ถูกต้อง ว่ามีค่า BMI เป็น “อ้วน” จำนวน 1 คน และ “เริ่มอ้วน” จำนวน 2 คน เป็นต้น

ในทางกลับกัน จากภาพที่ 10 เมื่อเราใช้วิธีการ SVM Model ในรูปแบบ Multiclass เพื่ออธิบายถึงความแม่นยำของข้อมูล จากตารางดังกล่าวจะเห็นว่า เมื่อมองในแนวทแยงของข้อมูลซึ่งแสดงถึงความถูกต้องของการทำนาย จะพบว่า การทำลายมีความถูกต้องมากกว่า อาทิเช่น จากข้อมูล 89 คนที่ค่าจริง “ปกติ” ในการทำนายทำนายว่า ปกติ ได้ถูกต้องจำนวน 89 คน ทำนายได้ถูกทั้งหมด เป็นต้น



ภาพที่ 10 แสดงถึงผลการทำนาย(Prediction) ด้วยวิธีการ Decision Tree Model



ภาพที่ 11 แสดงถึงผลการทำนาย(Prediction) ด้วยวิธีการ SVM Model

ข้อเสนอแนะ

1. ข้อเสนอแนะสำหรับการนำผลการวิจัยไปใช้

ในการนำเสนอให้กับมหาวิทยาลัยราชภัฏในแต่ละแห่ง พบว่า ข้อมูลท้องถิ่นที่เกิดขึ้นในอดีตมีความหลากหลาย หากสามารถการใช้มาตรฐานของข้อมูลท้องถิ่นเดียวกัน จะตอบโจทย์ และขยายขีดความสามารถในการกำกับติดตาม คุณภาพชีวิต ของชุมชนในท้องถิ่น ที่ได้รับการส่งเสริมสนับสนุนจาก กิจกรรมต่าง ๆ ของทางมหาวิทยาลัยราชภัฏ นอกจากนี้ การนำกลุ่มข้อมูลราชภัฏ โดยเฉพาะอย่างยิ่งข้อมูลท้องถิ่น มาใช้และแบ่งปันร่วมกันของมหาวิทยาลัยราชภัฏ จะเป็นการประสานความร่วมมือระหว่างมหาวิทยาลัยราชภัฏได้อย่างมีประสิทธิภาพ

นอกจากนี้การนำระบบปัญญาประดิษฐ์มาใช้ในการช่วยในการทำนายข้อมูล เพื่อให้ได้ข้อมูลที่มีความถูกต้องและใช้ประโยชน์ได้อย่างเต็มที่ ซึ่งหลักการนี้จะได้ผลลัพธ์ที่แตกต่างจากการที่ไม่ได้นำมาประยุกต์ใช้กับกลุ่มข้อมูลราชภัฏ ในการสนับสนุนทิศทางกิจกรรม โครงการต่าง ๆ ได้

2. ข้อเสนอแนะสำหรับการทำวิจัยครั้งต่อไป

อย่างไรก็ตาม การดำเนินการดังกล่าวจำเป็นจะต้อง จะเก็บข้อมูลอย่างต่อเนื่องเพื่อจะให้ทราบถึงระยะเวลาและคาดการณ์แนวโน้มที่จะเกิดขึ้นในอนาคต เป็นเรื่องที่ต้องการร่วมมือจากทุกมหาวิทยาลัยราชภัฏ เพื่อผลักดันให้ เป็นข้อมูลท้องถิ่นขนาดใหญ่ (Big Data) ที่มีความสมบูรณ์ครอบคลุมทุกพื้นที่ในการให้บริการทั้งประเทศ เพื่อใช้เป็นข้อมูลที่สำคัญในการขับเคลื่อนบริบท ของการเป็นมหาวิทยาลัย ในการพัฒนาท้องถิ่น เพิ่มขีดความสามารถทางการแข่งขันได้เป็นอย่างดี

ดังนั้น การติดตามของช่วงอายุของประชากร เป็นวิ่งที่จะใช้ในการกำหนดจำนวนของผู้ที่จะปรับเปลี่ยนไปในสถานะต่าง ๆ ได้อย่างชัดเจน เป็นสิ่งบ่งชี้ความสำเร็จของโครงการหรือกิจกรรมที่ทางมหาวิทยาลัยราชภัฏลงพื้นที่ไปสนับสนุนส่งเสริม เพื่อให้มีคุณภาพชีวิตที่ดีและสำเร็จตรงตามวัตถุประสงค์ของการพัฒนาท้องถิ่นได้อย่างชัดเจน อาทิ การติดตามสภาวะเริ่มอ้วนของประชากรกลุ่มเยาวชนหรือประชากรที่มีสถานะอ้วนในช่วงอายุที่จะเข้าสู่วันรุ่น หรือ เข้าสู่วันกลางคน หรือแม้ประชากรที่จะเข้าสู้ผู้สูงวัย

เอกสารอ้างอิง

Broda, M. D., Bogenschutz, M., Dinora, P., Prohn, S. M., Lineberry, S., & Ross, E. (2021). Using Machine Learning to Predict Patterns of Employment and Day Program Participation. **American Journal on Intellectual and Developmental Disabilities, 126(6),** 477–491.

Jena, K. K., Bhoi, S. K., Mohapatra, D., Mallick, C., & Swain, P. (2021). Rice Disease Classification Using Supervised Machine Learning Approach. In **Proceedings of the 5th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2021,** 328–333.

Lek, D., Haveman-Nies, A., Bezem, J., Zainalabedin, S., Schetters-Mouwen, S., Saat, J., Gort, G., Roovers, L., & van Setten, P. (2021). Two-year effects of the community-based overweight and obesity intervention program Gezond Onderweg! (GO!) in children and adolescents living in a low socioeconomic status and multi-ethnic district on Body Mass Index-Standard Deviation Score and quality of life. **EClinicalMedicine, 42.**

Myers, S., Govindarajulu, U., Joseph, M. A., & Landsbergis, P. (2021). Work Characteristics, Body Mass Index, and Risk of Obesity: The National Quality of Work Life Survey. **Annals of Work Exposures and Health, 65(3),** 291–306.

Nanglia, P., Kumar, S., Mahajan, A. N., Singh, P., & Rathee, D. (2021). A hybrid algorithm for lung cancer classification using SVM and Neural Networks. **ICT Express, 7(3),** 335–341.

Saetang, W., Tangwannawit, S., & Jensuttiwetchakul, T. (2021). การศึกษาการยอมรับเทคโนโลยีข้อมูลขนาดใหญ่ในประเทศไทย: มุมมองขององค์กร (A STUDY OF BIG DATA TECHNOLOGY ADOPTION IN THAILAND: ORGANIZATIONAL PERSPECTIVE). **วารสารมหาวิทยาลัยศรีนครินทรวิโรฒ สาขาวิทยาศาสตร์และเทคโนโลยี, 13(25, January-June),** 110–122.

Simmonds, M., Llewellyn, A., Owen, C. G., & Woolacott, N. (2016). Predicting adult obesity from childhood obesity: a systematic review and meta-analysis. **Obesity Reviews, 17(2),** 95–107.

Sulochana, V., Shanthini, B., & Harinath, K. (2022). Fast Intraprediction Algorithm for HEVC Based on Machine Learning Classification Technique. In **IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics, ICDCECE 2022.**

Tamayo, M. C., Dobbs, P. D., & Pincu, Y. (2021). Family-Centered Interventions for Treatment and Prevention of Childhood Obesity in Hispanic Families: A Systematic Review. **Journal of Community Health, 46(3),** 635–643.

Zhao, H., Zuo, X., & Xie, Y. (2022). Customer Churn Prediction by Classification Models in Machine Learning. In **2022 9th International Conference on Electrical and Electronics Engineering, ICEEE 2022,** 399–407.