



แนวโน้มการเลือกแผนการศึกษาต่อโดยใช้การเรียนรู้ของเครื่อง ด้วยการเปรียบเทียบผลระหว่าง ARTIFICIAL NEURAL NETWORK และ SUPPORT VECTOR MACHINE

TREND OF SELECTION FOR STUDY PLANNING USING MACHINE LEARNING BY COMPARING THE RESULTS BETWEEN ARTIFICIAL NEURAL NETWORK AND SUPPORT VECTOR MACHINE

Received: 24 October 2023 Revised: 19 December 2023 Accepted: 22 December 2023

ฐิติชัย รักบำรุง, คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา
พัสกร แนวปรานี, โรงเรียนสาธิต “พิบูลบำเพ็ญ” มหาวิทยาลัยบูรพา
ชลธิชา ฐิริปาณิก, โรงเรียนสาธิต “พิบูลบำเพ็ญ” มหาวิทยาลัยบูรพา
พรสุตา ชูพันธ์, โรงเรียนสาธิต “พิบูลบำเพ็ญ” มหาวิทยาลัยบูรพา

บทคัดย่อ รายงานการวิจัยฉบับนี้เป็นการหาแนวโน้มการเลือกแผนการศึกษาต่อระดับมัธยมศึกษาตอนปลาย ระหว่างโปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์ กับโปรแกรมเน้นความสามารถทางภาษา โดยใช้การเรียนรู้ของเครื่องมาช่วยวิเคราะห์ข้อมูลแบบ Supervised Learning เปรียบเทียบ 2 เทคนิค ได้แก่ เทคนิค Artificial Neural Network และเทคนิค Support Vector Machine ด้วยจำนวนชุดข้อมูลตัวอย่าง 908 ชุด วัตถุประสงค์การวิจัย 1) เพื่อสร้างแบบจำลองการวิเคราะห์แนวโน้มการเลือกแผนการศึกษาต่อระหว่างโปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์กับโปรแกรมเน้นความสามารถทางภาษา 2) เพื่อเปรียบเทียบประสิทธิภาพแบบจำลองการเลือกแผนการศึกษาต่อระหว่าง เทคนิค Artificial Neural Network และเทคนิค Support Vector Machine พบว่า เมื่อนำข้อมูลมาสร้างแบบจำลองหาร้อยละค่าความถูกต้อง (accuracy) โดยการแบ่งข้อมูลชุดเรียนรู้ (training set) ร้อยละ 80 และข้อมูลชุดทดสอบ (testing set) ร้อยละ 20 การวิเคราะห์ค่าความถูกต้อง (accuracy) พบว่า 1) เทคนิค Artificial Neural Network มีค่าความถูกต้องที่ร้อยละ 80 2) เทคนิค Support Vector Machine มีค่าความถูกต้องที่ร้อยละ 85

คำสำคัญ: Machine Learning, Artificial Neural Network, Support Vector Machine

Abstract This research is to finding trends for decision in the selection of high school study plans between sciences and mathematics plan and arts plan by using machine learning to help analyze data using Supervised Learning by comparing 2 techniques;



Artificial Neural Network and Support Vector Machine with a total of 908 sample data sets. The objectives of research were 1) to create a model of trends analyze for decision to select high school study plans between sciences and mathematics plan and arts plan 2) to compare the performance of the selection of high school study plans model by Artificial Neural Network and Support Vector Machine techniques. It was found that when the data were used to find the percentage of accuracy by dividing of the training set; the training set was 80 percent and the testing set was 20 percent. The accuracy analyzing was as follows: 1) Artificial Neural Network 80 percent 2) Support Vector Machine 85 percent.

Keywords: Machine Learning, Artificial Neural Network, Support Vector Machine

บทนำ

ความก้าวหน้าทางด้านวิทยาศาสตร์และเทคโนโลยีก่อให้เกิดการเปลี่ยนแปลงหลากหลายมิติ ทั้งทางด้านสังคม เศรษฐกิจ การติดต่อสื่อสารทำได้อย่างรวดเร็วและกว้างขวางทั่วโลก ก่อให้เกิดการเรียนรู้ วัฒนธรรมภาษาที่แตกต่างหลากหลายอย่างมาก การเปลี่ยนสิ่งเหล่านี้เกิดขึ้นอยู่ตลอดเวลาในลักษณะที่เป็นพลวัต (Dynamic) ส่งผลให้สถาบันการศึกษาต้องปรับตัวด้วยการออกแบบและพัฒนาหลักสูตร ที่ตอบสนองความต้องการเพื่อเสริมสร้างผู้เรียนให้ทันต่อแข่งขัน และการปรับกระบวนทัศน์ใหม่ (New Paradigm) เพื่อให้ก้าวทันกับการเปลี่ยนแปลงของสังคมโลก

โรงเรียนสาธิต “พิบูลบำเพ็ญ” มหาวิทยาลัยบูรพา ได้ออกแบบและพัฒนาโปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์ และโปรแกรมเน้นความสามารถทางภาษา โดยโปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์ ให้ความสำคัญในการจัดกิจกรรมการเรียนรู้ที่เน้นด้าน วิทยาศาสตร์และคณิตศาสตร์ ดูแลและจัดกิจกรรมสำหรับผู้เรียนโดยอาจารย์ที่มีความรู้ความสามารถและความเชี่ยวชาญเฉพาะด้าน เพื่อส่งเสริมการพัฒนาผู้เรียนให้มีทักษะด้านการคิด วิเคราะห์ และแก้ปัญหา อย่างเป็นระบบ มีทักษะกระบวนการทางวิทยาศาสตร์ขั้นสูง เพื่อนำไปสู่การสร้างองค์ความรู้ที่เหมาะสม นอกจากนี้โรงเรียนสาธิต “พิบูลบำเพ็ญ” มหาวิทยาลัยบูรพา ยังมีโปรแกรมเน้นความสามารถทางภาษา ที่มีการจัดกิจกรรมการเรียนรู้ครอบคลุมและหลากหลาย เน้นการทำกิจกรรมด้านภาษาต่างประเทศเพื่อให้ผู้เรียนสามารถพัฒนาทักษะภาษาที่ตนสนใจ โดยได้รับความรู้โดยตรงจากอาจารย์เจ้าของภาษา ได้แก่ ภาษาอังกฤษ จีน ญี่ปุ่น ฝรั่งเศส เพื่อสร้างเสริมให้ผู้เรียนได้ศึกษาเรียนรู้จากประสบการณ์จริงทั้งในและนอกห้องเรียนรวมทั้งการได้ไปฝึกภาษาโดยตรงในต่างประเทศที่ผู้เรียนสนใจ ทำให้ผู้เรียนมีทักษะทางภาษา ด้านการฟัง พูด อ่าน เขียน และสร้างทัศนคติที่ดีต่อภาษาต่างประเทศ (โรงเรียนสาธิต “พิบูลบำเพ็ญ” มหาวิทยาลัยบูรพา, 2566)

การศึกษาต่อชั้นมัธยมศึกษาตอนปลาย ผู้เรียนจำเป็นต้องตัดสินใจเลือกแผนการศึกษาเพื่อให้เกิดความเหมาะสมกับความถนัดและความสามารถที่มี โดยเฉพาะอย่างยิ่งผู้เรียนที่มีความประสงค์จะศึกษาต่อ



ในโปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์ และโปรแกรมเน้นความสามารถทางภาษา การเลือกแผนการศึกษาต่อที่เหมาะสมเป็นเรื่องยากในการตัดสินใจเนื่องจากผู้เรียนอายุยังน้อยยังขาดประสบการณ์และการทำความเข้าใจต่อตนเองอาจไม่ทราบรายละเอียดแผนการเรียนหรือสิ่งที่ต้องการศึกษาดีพอ จึงอาจส่งผลต่อการตัดสินใจเลือกโปรแกรมการเรียนที่จะส่งผลต่ออนาคตทำให้เกิดความไม่แน่ใจหรือเลือกโปรแกรมการเรียนไม่เหมาะสมตามศักยภาพที่ตนมี แนวทางหนึ่งในการเลือกแผนการเรียน คือ การนำข้อมูลผู้เรียนมาวิเคราะห์หาแนวโน้มการศึกษาต่อเพื่อประกอบการตัดสินใจ

การเรียนรู้ของเครื่อง เป็นรูปแบบหนึ่งที่ได้รับการนิยมนำมาช่วยประมวลผลข้อมูล โดยการนำข้อมูลที่เกี่ยวข้องกับเรื่องที่ต้องการจะประมวลผลมาสร้างแบบจำลองมีความแม่นยำในการทำนายสูงในการทำนายผล จึงนิยมนำมาประยุกต์ใช้ได้หลากหลายงานเป็นที่ยอมรับระดับสากล (Witten and Frank, 2005) ได้มีตัวอย่างที่ได้นำการเรียนรู้ของเครื่องมาใช้งานวิจัย สำหรับการแนะนำการเรียนของผู้เรียน โดยนำวิธีการนี้มาใช้ในการวิเคราะห์เพื่อการทำนายแนวโน้มการเรียนที่มีความเหมาะสมกับผู้เรียน (Aly et al., 2013) การเรียนรู้ของเครื่อง สามารถทำการวิเคราะห์ผ่านโปรแกรมและภาษาคอมพิวเตอร์ได้หลากหลายรูปแบบ เช่น Scikit-learn, Rapidminer, Weka, Tensorflow, Python เป็นต้น การเลือกใช้งานภาษาคอมพิวเตอร์ และโปรแกรมเหล่านี้จะขึ้นอยู่กับลักษณะชุดข้อมูล จำนวน ความสามารถในการเขียนชุดคำสั่ง การแสดงผลของข้อมูล ในงานวิจัยนี้เลือก Scikit-learn (Géron, 2017) ซึ่งเป็นโอเพ่นซอร์ส (open source) ถูกใช้บนภาษา Python ที่มีความสามารถในการประมวลผล ควบคุม สั่งการข้อมูลได้อย่างอิสระ (Drake, 2003) โดยมีความสามารถในการประมวลผลการทำงานผ่านชุดคำสั่งจากภาษา Python ผ่านโปรแกรมที่ป้อนคำสั่งให้

งานวิจัยนี้เลือกใช้การเรียนรู้ของเครื่องในแบบ Supervised Learning ประกอบด้วย 2 เทคนิค คือ 1) Artificial Neural Network 2) Support Vector Machine ด้วยเหตุผลที่ Artificial Neural Network เป็นเทคนิคที่สามารถหาความสัมพันธ์เมื่อมีอินพุตโดเมนหรือฟังก์ชันและได้ผลลัพธ์ออกมาโครงข่ายประสาทเทียมโดยปกติจะเป็นเลเยอร์โหนดของกราฟสต็อคคล้องกับเซลล์ประสาทและการเชื่อมโยงไปยังซินแนปส์ (Adcock et al., 2015) นิวรอนเน็ตเวิร์กจะทำให้เราทราบว่าฟังก์ชันที่เหมาะสมกับโดเมนหรือ input คืออะไร และ Support Vector Machine เป็นวิธีที่ได้รับความนิยมในการทำข้อมูลลักษณะแยกประเภทข้อมูลได้ดี (Pupale, 2018) ทั้งมียืดหยุ่นในการวิเคราะห์ข้อมูล โดยเฉพาะอย่างยิ่งเมื่อข้อมูลมีความซับซ้อนหลายตัวแปร ผู้วิจัยจึงได้ทำการวิจัยนี้

วัตถุประสงค์ของการวิจัย

- 1) เพื่อสร้างแบบจำลองการวิเคราะห์แนวโน้มการเลือกแผนการศึกษาต่อระหว่างโปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์กับโปรแกรมเน้นความสามารถทางภาษา
- 2) เพื่อเปรียบเทียบประสิทธิภาพแบบจำลองการเลือกแผนการศึกษาต่อระหว่าง เทคนิค Artificial Neural Network และเทคนิค Support Vector Machine

ขอบเขตการวิจัย

ขอบเขตด้านข้อมูล

ข้อมูลที่ใช้ในงานวิจัยนี้เป็นข้อมูลของผู้สำเร็จชั้นมัธยมศึกษาปีที่ 4-6 ปีการศึกษา 2554 - 2564 จำนวน 908 ชุดข้อมูล ประกอบด้วย Attributes ดังต่อไปนี้

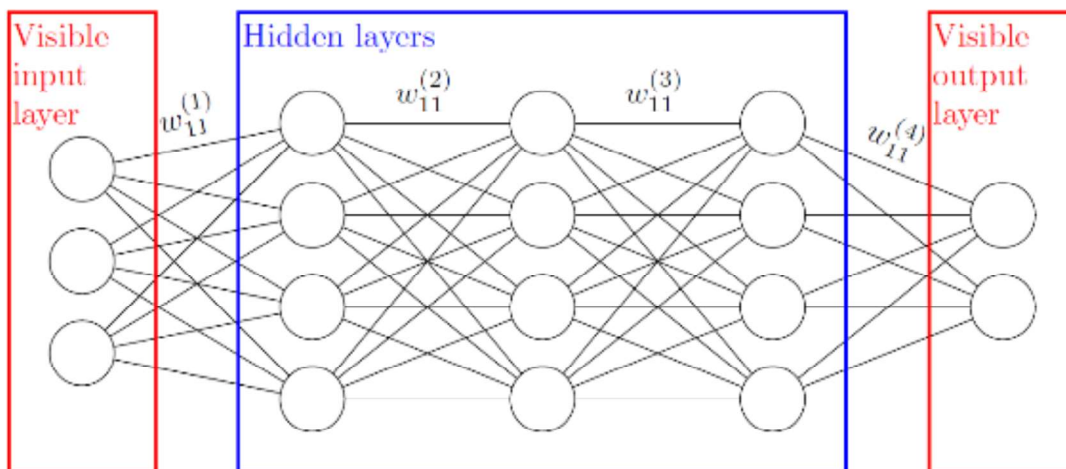
- 1) เพศ
- 2) ภูมิำเนา
- 3) อาชีพของผู้ปกครอง
- 4) รายได้ผู้ปกครอง
- 5) แผนการเรียน
- 6) วิชาคณิตศาสตร์
- 7) วิชาวิทยาศาสตร์
- 8) วิชาภาษาอังกฤษ
- 9) วิชาภาษาไทย
- 10) วิชาสังคม
- 11) วิชาศิลปะ
- 12) วิชาสุขศึกษา
- 13) วิชาประวัติศาสตร์
- 14) วิชาพลศึกษา
- 15) วิชาการงานอาชีพและเทคโนโลยี
- 16) เกรดเฉลี่ยมัธยมศึกษาปีที่ 3
- 17) เกรดเฉลี่ยมัธยมศึกษาปีที่ 6

ขอบเขตด้านอัลกอริทึมที่นำมาใช้ในการสร้างแบบจำลองในการทำนายแผนการศึกษาต่อ

- 1) Artificial Neural Network
- 2) Support Vector Machine

เอกสารที่เกี่ยวข้อง

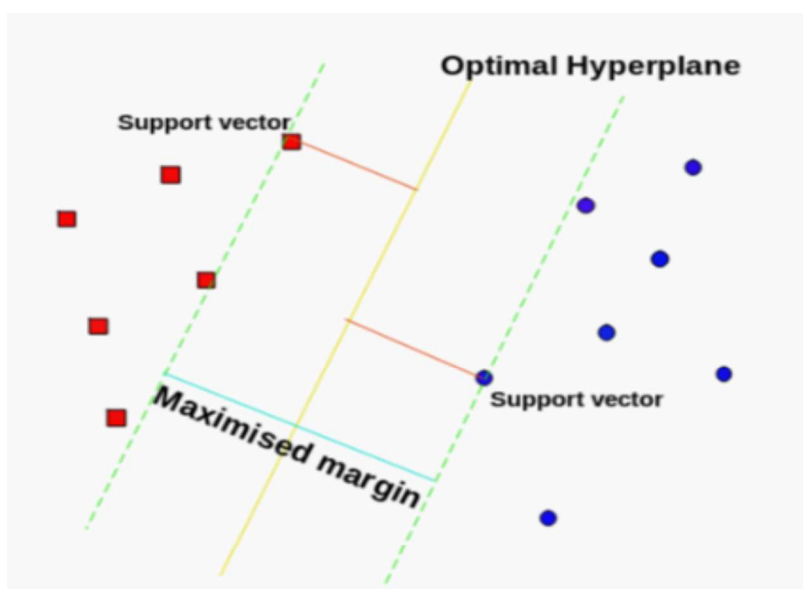
โครงข่ายประสาทเทียม (Artificial Neural Network) คือเป็นเทคนิคที่จำลองระบบประสาทของสิ่งมีชีวิตเพื่อใช้ในการวิเคราะห์อยู่ในรูปแบบสมการคณิตศาสตร์หรือแบบจำลองที่คอมพิวเตอร์สามารถจำลองการทำงานผลและมีค่าการคำนวณออกมาคล้ายกระบวนการทางโครงข่ายประสาท



ภาพที่ 1 โครงข่ายประสาทเทียม (Adcock, 2015)

จากภาพประกอบเป็นลักษณะโครงข่ายประสาทเทียมแบบที่มีเลเยอร์เรียงซ้อนอยู่ 3 ชั้น แต่ละชั้นจะถูกระบุในลักษณะเวกเตอร์ของส่วนประกอบแบบไบนารีโดยมีขอบระหว่างเวกเตอร์กำหนดเป็นเมทริกซ์ของค่าน้ำหนัก พื้นที่การกำหนดค่าของกราฟที่ได้รับจากการแจกแจง Gibbs ด้วย Ising-spin Hamiltonian (Adcock, 2015)

ซัพพอร์ตเวกเตอร์แมชชีน (support vector machine) เป็นแบบจำลองการทำนายเชิงเส้นเพื่อการจำแนกประเภทและปัญหา โดยเป็นอัลกอริทึมที่สร้างเส้นที่อาจเรียกว่าไฮเปอร์เพลนโดยการสร้างเส้นไฮเปอร์เพลนนี้เพื่อใช้เป็นตัวแยกข้อมูลออกเป็นคลาสระหว่างข้อมูลที่จะแบ่งออกเป็นสองคลาส โดยมีจุดมุ่งหมายหรือเป้าหมายหลักในการที่จะใช้เส้นที่แบ่งมีความกว้างมากที่สุด แต่ถ้าเส้นแคบไปก็ยังไม่ยอมให้มีข้อมูลอยู่ระหว่างเส้นแบ่ง ดังภาพที่ 2 (Pupale, 2018)



ภาพที่ 2 ซัพพอร์ตเวกเตอร์แมชชีน (support vector machine) (Pupale, 2018)



Scikit-Learn เป็นโมดูล แบบโอเพนซอร์ซ (open source) New BSD License ที่ใช้ภาษา Python, Numpy และ Scipy โดยมีจุดเด่น คือฟังก์ชันในการใช้งานเพื่อการแบ่งประเภทข้อมูล การแบ่งกลุ่มชุดข้อมูล การวิเคราะห์ที่มีความหลากหลายรูปแบบใช้งานได้อย่างอิสระ และรวมอัลกอริธึมการเรียนรู้ของเครื่องที่ครอบคลุมการใช้งานมากกว่าการเลือกใช้โปรแกรมสำเร็จรูปในการทำนายผล เหมาะสำหรับทั้งผู้ที่มีความเชี่ยวชาญในการใช้ภาษาโปรแกรมเมอร์รวมทั้งระดับผู้เริ่มต้นใช้งานก็สามารถเรียนรู้การใช้งานได้ง่ายสำหรับการใช้การเรียนรู้ของเครื่อง (Pedregosa et al, 2011)

วิธีการวิจัย

งานวิจัยฉบับนี้เป็นการหาแนวโน้มการเลือกแผนการศึกษาต่อระดับมัธยมศึกษาตอนปลาย ระหว่างโปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์ กับโปรแกรมเน้นความสามารถทางภาษา โดยใช้การเรียนรู้ของเครื่องมาช่วยวิเคราะห์ข้อมูลแบบ Supervised Learning เปรียบเทียบ 2 เทคนิคได้แก่ Artificial Neural Network และ Support Vector Machine สำหรับนักเรียนมัธยมศึกษาตอนต้นโรงเรียนสาธิต “พิบูลบำเพ็ญ” มหาวิทยาลัยบูรพา โดยมีขั้นตอนดังนี้ (Chapman et al., 2000 and Witten and Frank, 2005) 1) การเลือกข้อมูล (data selection) 2) การกลั่นกรองข้อมูล (data cleansing) 3) การแปลงข้อมูล (data transformation) 4) การสร้างแบบจำลอง (model) ทดสอบและเปรียบเทียบประสิทธิภาพเทคนิค และ 5) สรุปผลการวิจัย โดยการวิจัยเริ่มต้นด้วยการเตรียมข้อมูล (data preparation) ดังนี้

1. การเลือกข้อมูล (Data Selection) นำข้อมูลที่ใช้ในการศึกษาเพื่อทำการวิเคราะห์หาแนวโน้มการเลือกแผนการศึกษาต่อระดับมัธยมศึกษาตอนปลาย ระหว่างโปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์ กับโปรแกรมเน้นความสามารถทางภาษา จำนวน 908 ชุดข้อมูล มีจำนวนแอททริบิวต์ทั้งหมด 70 แอททริบิวต์ ได้แก่ เพศ ระดับผลการเรียนรายวิชาหลัก 10 วิชา ผลคะแนนเรียนสะสมเฉลี่ย

2. การกลั่นกรองข้อมูล (Data Cleansing) นำข้อมูลระดับผลการเรียนรายวิชาหลัก 10 วิชา ประกอบด้วย 1) ภาษาไทย 2) คณิตศาสตร์ 3) วิทยาศาสตร์ 4) สังคมศึกษา 5) ประวัติศาสตร์ 6) สุขศึกษา 7) ศิลปะ 8) ภาษาต่างประเทศ 9) การงานอาชีพและเทคโนโลยี 10) พละศึกษา มาจัดกระทำให้อยู่ในรูปแบบคะแนนเฉลี่ยรวม จำนวน 6 ภาคการศึกษา และหาค่าเกรดเฉลี่ยตลอดหลักสูตรชั้นมัธยมศึกษาตอนต้น

3. การแปลงข้อมูล (Data Transformation) ผู้วิจัยทำการแปลงข้อมูลเพื่อให้เหมาะสมที่จะนำมาวิเคราะห์และสร้างแบบจำลองการทำนาย ดังนี้ 1) ข้อมูลผลการเรียนรายวิชาหลัก 10 วิชา ที่ถูกจัดกระทำในรูปแบบคะแนนเฉลี่ยรวม จำนวน 6 ภาคการศึกษา นำมาแบ่งเป็น 3 ช่วง โดยกำหนดระดับได้แก่ระดับคะแนน ช่วงที่ 1 เท่ากับ Fair มีค่าเท่ากับ 1 ช่วงที่ 2 เท่ากับ Good มีค่าเท่ากับ 2 ช่วงที่ 3 เท่ากับ Very Good มีค่าเท่ากับ 3 2) แอททริบิวต์เพศ ให้ค่าเพศชายเท่ากับ 1 และเพศหญิงเท่ากับ 2 3) เกรดเฉลี่ยการศึกษา โดยกำหนดระดับคะแนนช่วงที่ 1 เท่ากับ Fair มีค่าเท่ากับ 1 ช่วงที่ 2 เท่ากับ Good มีค่าเท่ากับ 2



ช่วงที่ 3 เท่ากับ Very Good มีค่าเท่ากับ 3 4) กำหนดให้โปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์เป็น 1 และโปรแกรมเน้นความสามารถทางภาษาเป็น 2

4. การสร้างแบบจำลอง (Model) ใช้ Scikit - learn ด้วยภาษา Python มีการทำงานแบบการแปลชุดคำสั่งที่ละบรรทัด เลือกใช้ Machine Learning ในแบบ Supervised Learning

5. การแปรผลและการประเมินผล (Interpretation Evaluation) นำเสนอผลการทดสอบประสิทธิภาพแบบจำลองทั้ง 2 วิธี 1) Artificial Neural Network 2) Support Vector Machine แสดงค่าต่อไปนี้

1) ค่าความแม่นยำ (Precision) แทนค่าในสมการ $Precision = TP / (TP + FP)$ ค่า TP หมายถึง ข้อมูลที่ทำนายแล้วถูกต้องเมื่อเทียบกับคำตอบ และ FP หมายถึงข้อมูลที่อยู่ในคำตอบแต่ไม่มีในการทำนาย

2) ค่าโอกาสการเรียกซ้ำ (Recall) แสดงค่าจากตารางด้วยค่าร้อยละตามลำดับในสมการ $Recall = TP / (TP + FN)$ ค่า FN หมายถึง ข้อมูลที่ทำนายแล้วไม่ถูกต้องเมื่อเทียบกับคำตอบ

3) F1 - Score เป็นการตรวจสอบความแม่นยำของแบบจำลอง คือ ค่าเฉลี่ยแบบ Harmonic Mean ระหว่าง Precision และ Recall เป็น Single Metric ที่วัดความสามารถของโมเดลแทนค่าในสมการ $F1 = 2 * ((Precision * Recall) / (Precision + Recall))$

4) จำนวนครั้งที่เกิดการเลือก (Support) ข้อมูล

5) ค่าร้อยละความถูกต้อง (Accuracy) แสดงค่าที่โมเดลทำนายถูกทั้งหมดเป็นค่าร้อยละ

เครื่องมือในการวิจัย

เครื่องมือสร้างแบบจำลองด้านซอฟต์แวร์และโปรแกรม (Software and Program) มีดังนี้

- 1) Anaconda
- 2) Bundle Library Scikit - Learn
- 3) Library Numpy
- 4) Library Pandas
- 5) Library Matplotlib
- 6) Jupyter Notebook (IDE : Integrated Development Environment)

เครื่องมือสร้างแบบจำลองด้านฮาร์ดแวร์ (Hardware) มีดังนี้

- 1) เครื่องคอมพิวเตอร์เพื่อการประมวลผลข้อมูล
- 2) การ์ดจอแยกอิสระ (Dedicated Graphics Cards)

ผลการวิจัย

จากแบบจำลองด้วยเทคนิค Artificial Neural Network ที่ได้แบ่งข้อมูลชุดเรียนรู้ (training set) ร้อยละ 80 และข้อมูลชุดทดสอบ (testing set) ร้อยละ 20 โดยโปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์แทนด้วยหมายเลข 1 และโปรแกรมเน้นความสามารถทางภาษาแทนด้วยหมายเลข 2

**ตารางที่ 1** ผลการทดสอบแบบจำลองด้วยเทคนิค Artificial Neural Network

Artificial Neural Network	N = 908			
	precision	recall	F1 - score	support
1	0.87	0.81	0.82	60
2	0.70	0.80	0.77	31
accuracy			0.80	91

จากตารางที่ 1 ผลการทดสอบแบบจำลองด้วยเทคนิค Artificial Neural Network มีค่าความแม่นยำ (precision) ของการเลือกระหว่างโปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์ กับโปรแกรมเน้นความสามารถทางภาษา มีค่าร้อยละ 87 และร้อยละ 70 ตามลำดับ ค่าโอกาสการเรียกซ้ำ (recall) โปรแกรมการเรียนรู้ที่ 1 และโปรแกรมการเรียนรู้ที่ 2 ที่ร้อยละ 81 และร้อยละ 80 ตามลำดับ ผลที่ได้จากแบบจำลองด้วยเทคนิค Artificial Neural Network จะใช้การตรวจสอบ แบบ F1 - score เพื่อตรวจสอบความแม่นยำของแบบจำลอง และจำนวนครั้งที่เกิดการเลือกโปรแกรมการเรียนรู้ที่ 1 หรือโปรแกรมการเรียนรู้ที่ 2 (support) อยู่ที่ร้อยละ 91 โดยค่าร้อยละความถูกต้อง (accuracy) ของเทคนิค Artificial Neural Network มีความถูกต้องที่ร้อยละ 80

จากแบบจำลองด้วยเทคนิค Support Vector Machine ที่ได้แบ่งข้อมูลชุดเรียนรู้ (training set) ร้อยละ 80 และข้อมูลชุดทดสอบ (testing set) ร้อยละ 20 โดยโปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์แทนด้วยหมายเลข 1 และโปรแกรมเน้นความสามารถทางภาษาแทนด้วยหมายเลข 2

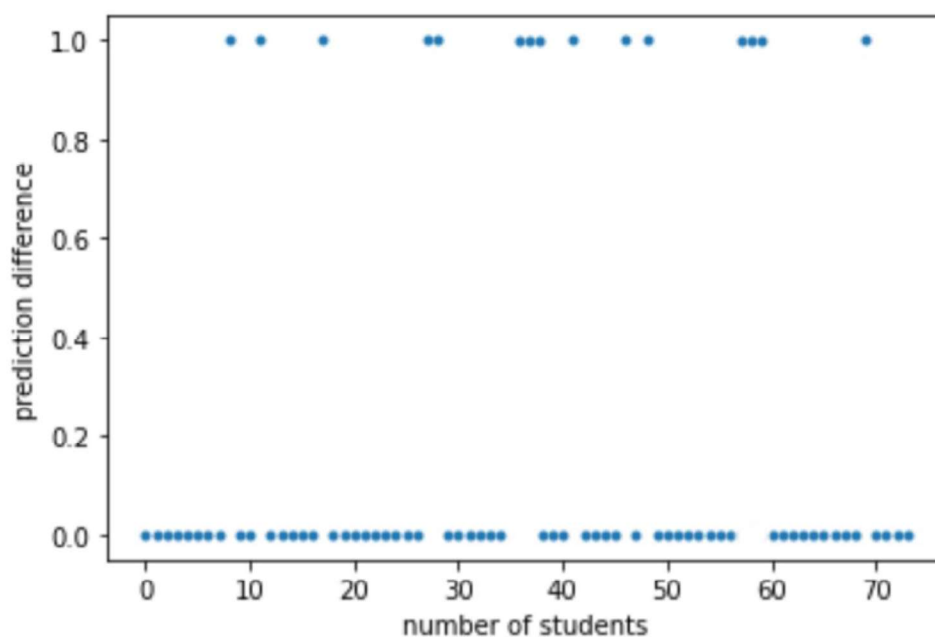
ตารางที่ 2 ผลการทดสอบแบบจำลองด้วยเทคนิค Support Vector Machine

Support Vector Machine	N = 908			
	precision	recall	F1 - score	support
1	0.88	0.90	0.89	51
2	0.77	0.74	0.76	23
accuracy			0.85	74

จากตารางที่ 2 ผลการทดสอบแบบจำลองด้วยเทคนิค Support Vector Machine มีค่าความแม่นยำ (precision) ของการเลือกระหว่างโปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์ กับโปรแกรมเน้นความสามารถทางภาษา มีค่าร้อยละ 88 และร้อยละ 77 ตามลำดับ ค่าโอกาสการเรียกซ้ำ (recall) โปรแกรมการเรียนรู้ที่ 1 และโปรแกรมการเรียนรู้ที่ 2 ที่ร้อยละ 90 และร้อยละ 74 ตามลำดับ ผลที่ได้จากแบบจำลองด้วยเทคนิค Support Vector Machine จะใช้การตรวจสอบ แบบ F1 - score เพื่อตรวจสอบความแม่นยำของแบบจำลอง และจำนวนครั้งที่เกิดการเลือกโปรแกรมการเรียนรู้ที่ 1 หรือ



โปรแกรมการเรียนรู้ที่ 2 (support) อยู่ที่ร้อยละ 74 โดยค่าร้อยละความถูกต้อง (accuracy) ของเทคนิค Support Vector Machine มีความถูกต้องที่ร้อยละ 85



ภาพที่ 3 ผลการทำนายด้วยเทคนิค Support Vector Machine

จากภาพอธิบายลักษณะการทำนายผลด้วยเทคนิค Support Vector Machine โดยการแบ่งข้อมูลชุดทดสอบ (testing set) ร้อยละ 20 เมื่อมีการทำนายถูกจะแสดง ณ ตำแหน่ง 0 และเมื่อทำนายผิด จะแสดง ณ ตำแหน่ง 1 เรียงตามลำดับจนครบชุดทดสอบ (testing set) จากภาพที่ 3 ค่าความถูกต้อง (accuracy) ที่ทำได้คือ ร้อยละ 85

ตารางที่ 3 ผลการเปรียบเทียบประสิทธิภาพแบบจำลอง

Algorithms	N = 908 Accuracy
Artificial Neural Network	80
Support Vector Machine	85

จากตารางที่ 3 ผลการเปรียบเทียบประสิทธิภาพแบบจำลองระหว่างเทคนิค Artificial Neural Network กับเทคนิค Support Vector Machine โดยแบบจำลองที่มีค่าความถูกต้อง (accuracy) มากที่สุด คือ แบบจำลองที่ใช้เทคนิค Support Vector Machine มีค่าความถูกต้อง (accuracy) ร้อยละ 85 และแบบจำลองที่ใช้เทคนิค Artificial Neural Network มีค่าความถูกต้อง (accuracy) ร้อยละ 80



อภิปรายผลและข้อเสนอแนะการวิจัย

การวิจัยครั้งนี้ใช้การเรียนรู้ของเครื่องมาสร้างแบบจำลองการทำนายและได้วิเคราะห์ข้อมูลผู้เรียนแบบ Supervised Learning เปรียบเทียบ 2 เทคนิคที่มีความแตกต่างกันเพื่อหาแนวโน้มการตัดสินใจเลือกโปรแกรมการเรียนมัธยมศึกษาตอนปลาย ระหว่างโปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์หรือโปรแกรมเน้นความสามารถทางภาษา โดยสองอัลกอริทึมนี้มีค่าความถูกต้อง (accuracy) แตกต่างกันถึงร้อยละ 5 เรียงตามลำดับดังนี้ 1) เทคนิค Artificial Neural Network ร้อยละ 80 2) เทคนิค Support Vector Machine มีค่าความถูกต้อง (accuracy) สูงสุด คือร้อยละ 85 สามารถอธิบายค่าที่สูงกว่าได้เนื่องจากเทคนิค Support Vector Machine (Pupale, 2018) เป็นวิธีที่เหมาะสมกับชุดข้อมูลที่ผู้วิจัยใช้โดยอัลกอริทึม Support Vector Machine นี้มีความสามารถแยกแยะข้อมูลในลักษณะหลายตัวแปร ซึ่งงานวิจัยนี้ใช้จำนวนแอตทริบิวต์ทั้งหมด 70 แอตทริบิวต์ ได้แก่ เพศ ระดับผลการเรียน รายวิชาหลัก 10 วิชา ผลคะแนนเรียนสะสมเฉลี่ย เป็นต้น และทำนายข้อมูลโดยเป็นการหาระยะห่างระหว่างกลุ่มข้อมูลที่สร้างเส้นแบ่งที่สามารถแยก (hyperplane) ที่ใกล้ที่สุดส่งผลให้เกิดความแม่นยำในการแยกชุดข้อมูลทำให้ค่าความแม่นยำของแบบจำลองสูง นอกจากนี้หากพิจารณาจากค่าความแม่นยำ (precision) ของการเลือกระหว่างโปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์ที่แทนด้วยหมายเลข 1 หรือโปรแกรมเน้นความสามารถทางภาษาแทนด้วยหมายเลข 2 จากตารางที่ 2 จะพบว่ามีค่าสูงร้อยละ 88 และร้อยละ 77 ตามลำดับ และค่าโอกาสการเรียกซ้ำ (recall) ที่ร้อยละ 90 และร้อยละ 74 ตามลำดับ จึงแสดงให้เห็นถึงความเหมาะสมในการใช้เทคนิค Support Vector Machine ในการหาแนวโน้มการศึกษาต่อมัธยมศึกษาตอนปลาย ซึ่งสอดคล้องกับ นนท์ บุญนิธิประเสริฐ และชัยพร เขมะภาคะพันธ์ ที่ได้นำเสนอวิธีการกรองสแปมในระบบส่งข้อความของโทรศัพท์เคลื่อนที่ ด้วยวิธีการกรองโดยใช้เทคนิค Support Vector Machine โดยในงานวิจัยดังกล่าวใช้หลักการของการกรองสแปม โดยการแบ่งคลาส 2 แบบ คือ YES หรือ NO ใช้ลักษณะเหมือนกันกับการเลือกระหว่างโปรแกรมเน้นความสามารถทางวิทยาศาสตร์และคณิตศาสตร์แทนด้วยหมายเลข 1 และโปรแกรมเน้นความสามารถทางภาษาแทนด้วยหมายเลข 2 โดยผลทดสอบด้วยวิธีการกรองแบบ Support Vector Machine มีความถูกต้องในการกรองข้อมูลสูงสุด (นนท์ บุญนิธิประเสริฐ และ ชัยพร เขมะภาคะพันธ์, 2552)

ข้อเสนอแนะการวิจัย

1) ในการสร้างแบบจำลองเพื่อช่วยในการหาแนวโน้มการศึกษาต่อต้องอาศัยตัวแปรจำนวนมาก ซึ่งการที่การเรียนรู้ของเครื่องที่ใช้เทคนิคแตกต่างกันจะเลือกตัวแปรไม่เหมือนกันในแต่ละเทคนิคทำให้ผู้วิจัยควรที่จะรวบรวมข้อมูลในทุกขอบเขตที่เป็นไปได้ในการจัดเก็บเพื่อเป็นการเพิ่มโอกาสที่จะเกิดความแม่นยำในการทำนายสูงสุด

2) ผู้วิจัยสามารถปรับแต่งวิธีการเรียนรู้ของเครื่องให้เหมาะสมเฉพาะชุดข้อมูลหรือใช้วิธีการอื่นเข้ามาร่วมในการวิเคราะห์ข้อมูลก่อนที่จะนำมาสร้างแบบจำลองการทำนายก็อาจจะทำให้มีความแม่นยำและถูกต้องมากยิ่งขึ้น



กิตติกรรมประกาศ

แนวโน้มการเลือกแผนการศึกษาต่อโดยใช้การเรียนรู้ของเครื่องด้วยการเปรียบเทียบผลระหว่าง Artificial Neural Network และ Support Vector Machine สำเร็จไปได้ด้วยดีอันเนื่องมาจากได้รับการสนับสนุนจาก คณะศึกษาศาสตร์ ภาควิชาคณิตกรรมและเทคโนโลยีการศึกษา โรงเรียนสาธิต “พิบูลบำเพ็ญ” มหาวิทยาลัยบูรพา ที่อำนวยความสะดวกในด้านขั้นตอนกระบวนการเอกสาร และขอขอบคุณ ผู้ทรงคุณวุฒิ ผู้เชี่ยวชาญด้านต่าง ๆ ที่กรุณาให้ความรู้ ความคิดเห็นในการปรับปรุงแก้ไขงานวิจัยนี้จนสำเร็จไปได้ด้วยดี ผู้วิจัยขอขอบคุณมา ณ โอกาสนี้

เอกสารอ้างอิง

- นนท์ บุญนิธิประเสริฐ และ ชัยพร เหมะภาคะพันธ์. (2552). *การกรองข้อความภาษาไทยและภาษาอังกฤษของบริการส่งข้อความสั้นบนเครือข่ายโทรศัพท์เคลื่อนที่*. Proceeding of The 7th National Conference on Computing and Information Technology (NCCIT2009), Bangkok, หน้า 34-39.
- โรงเรียนสาธิต “พิบูลบำเพ็ญ” มหาวิทยาลัยบูรพา. (2566). *หลักสูตรของโรงเรียน*. เข้าถึงได้จาก <http://www.st.buu.ac.th/html/index.php/th/abouts-th/course>
- Adcock, J.C. Allen, E. Day, M. Frick, S. Hinchliff, J. Johnson, M. Morley-Short, S. Pallister, S. Price, A.B. & Stanistic, S. (2015). *Advances in quantum machine learning*. Retrieved from <https://www.researchgate.net/publication/286513346>
- Aly, W. M., Hegazy. O. F., & Rashad, H. M. N. (2013). Automated student advisory using Machine Learning. *International Journal of Computer Applications*, 8(19), p.p 19-24.
- Chapman, P. Clinton, J. Kerber, R. Khabaza, T. Reinartz, T. Shearer, C. and Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*. SPSS Inc.
- Drake, Guido Van Rossum Fred L. (2003). *An introduction to Python*. Network Theory Limited.
- Géron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow*. CA: O'Reilly Media.
- Han, J. and Kamber, M. (2001). *Data mining concepts and techniques*. The Morgan Kaufmann Publishers.
- Jaroenpuntaruk, V. Wichadakul, D. (2015). Utilizing data mining techniques to forecast student academic achievement of Kasetsart University Laboratory School Kamphaeng Saen Campus Educational Research and Development Center. *Veridian E-J Sci Technol Silpakorn Univ*, 2(2), 1-17.



- Joyce, J. (2003). *Bayes' theorem*. Retrieved from <https://plato.stanford.edu/archives/spr2019/entries/bayes-theorem/>
- Mitchell, T.M. (1997). *Machine learning*. McGraw-hill.
- Pedregosa, F. Varoquaux, G. Gramfort, A. Michel, V. Thirion, B. Grisel, O. Blondel, M. Prettenhofer, P. Weiss, R. Dubourg, V. Vanderplas, J. Passos, A. Cournapeau, D. Brucher, M. Perrot, M. & Duchesnay, É. Scikit-learn. (2011). Machine learning in Python. *Journal of Machine Learning Research*. 12(85), p.p. 28252830.
- Pupale, R. (2018). *Support Vector Machines (SVM) - An Overview*. Retrieved from <https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989>
- Witten, I. H. and Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. The Morgan Kaufmann Publishers.