การศึกษาวิเคราะห์ข้อมูลผู้ที่ผ่านการคัดเลือกเข้าเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบก โดยใช้เทคนิคการจัดกลุ่มและแสดงผลด้วยกราฟดาต้าวิชวลไลเซชัน

A Study and Data Analytics for Selected AFAPS (Army Cadet) Candidates Using Classification Technique and Graph Data Visualization

พันโทหญิง รุ่งรัศมี สุวรรณวัฒนา[1]*
ผู้ช่วยศาสตราจารย์ ดร. สุรศักดิ์ มังสิงห์[2]

*[1]*กองวิชาวิศวกรรมไฟฟ้า ส่วนการศึกษา โรงเรียนนายร้อยพระจุลจอมเกล้า*
[2]สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยศรีปทุม
**Corresponding Author. E-mail: rungrasamee.su@crma.ac.th*
(Received: November 27, 2017; Revised: December 26, 2018; Accepted: June 6, 2018)

บทคัดย่อ : การศึกษาครั้งนี้ได้ออกแบบการศึกษาด้วยวิธีวิจัยแบบผสมวิธีจากเอกสาร (Document Research) และแบบการอธิบายต่อเนื่อง (Explanatory Sequential Mixed Methods) จากการศึกษาเอกสารและข้อมูล เพื่อนำเสนอตัวแทนข้อมูลด้วยกราฟดาต้าวิชวลไลเซชัน มุ่งเน้นที่การเลือกตัวอย่างแบบเฉพาะเจาะจงเป็นหลัก มีวัตถุประสงค์เพื่อศึกษาวิเคราะห์ความสัมพันธ์และการค้นพบรูปแบบของผู้ที่ผ่านการคัดเลือกเข้าเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบก ระหว่างปีการศึกษา 2554-2558 จำนวน 1,226 เรคอร์ด ผู้วิจัยได้เสนอแนวคิดในการวิเคราะห์ข้อมูลและจำแนกข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจ สถิติพื้นฐาน และแสดงผลด้วยกราฟดาต้าวิชวลไลเซชันในการเล่าเรื่องที่ผ่านมา และกำหนดปัจจัยพื้นฐานออกได้เป็น 7 ปัจจัยหลัก ได้แก่ อายุ เกรดเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3 วุฒิการศึกษาตอนสอบเข้า สถาบันการศึกษาเดิม จังหวัดสถานศึกษา ภูมิภาค สถานภาพครอบครัว โดยข้อมูลปัจจัยพื้นฐานเหล่านี้เป็นหนึ่งในองค์ประกอบที่ส่งผลให้ผ่านการคัดเลือกเข้าเป็นนักเรียนเตรียมทหาร ผลการศึกษาพบว่า จากการจำแนกตามปัจจัยอายุ พบว่า อายุ 17 ปี คือ ผู้ที่เข้ารับการศึกษามีจำนวนสูงสุดในแต่ละปีการศึกษา สามารถอนุมานได้ว่าเมื่อการสั่งสมประสบการณ์ ความรู้ ทักษะ ความสามารถในระดับหนึ่งแล้วจะสามารถผ่านการทดสอบได้สูง ด้วยการพยากรณ์ข้อมูลแนวโน้มจำนวนผู้ได้เข้ารับการศึกษา โดยใช้วิธีคาดคะเนแนวโน้มจากสมการแนวโน้มเชิงเส้น พบว่า มีจำนวนผู้เข้ารับการศึกษามีเกณฑ์เพิ่มขึ้น คิดเป็น 20.23 เปอร์เซ็นต์ ค่าเฉลี่ยของจำนวนผู้ที่ผ่านการสอบและเข้ารับการศึกษาที่ได้คือ 245 คน และค่าเบี่ยงเบนมาตรฐานข้อมูลคือ 1.79 เนื่องจากค่าเฉลี่ยข้อมูลมีความใกล้เคียงกัน ทั้งนี้ขึ้นอยู่กับปัจจัยภายนอกหรือความต้องการจากหน่วยขึ้นตรงร้องขอในแต่ละปีการศึกษา และการจำแนกตามปัจจัยเกรดเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3 โดยกำหนดให้ความสัมพันธ์แนวโน้มผลการเรียนแทนค่าด้วยฟังก์ชันเอกซ์โพเนนเชียล เมื่อนำข้อมูลเข้าโดยใช้เครื่องมือการทำเหมืองข้อมูลด้วย RapidMiner และนำผลที่ได้มาเปรียบเทียบกันพบว่า แนวโน้มของกลุ่มข้อมูลเกรดเฉลี่ยช่วง 3.50-4.00 อยู่ในระดับดีมาก มีประสิทธิภาพในการจำแนกข้อมูล และค่าเฉลี่ยที่ได้คือ 3.51 ค่าเบี่ยงเบนมาตรฐานข้อมูลคือ 0.40 อยู่ในเกณฑ์ที่ยอมรับได้ ผลการศึกษาทั้งสองนี้เป็นเพียงส่วนหนึ่งในการวิเคราะห์ข้อมูลประกอบการตัดสินใจ

และวางแผนนำเสนอในการจัดการข้อมูล และการแสดงผลด้วยกราฟดาต้าวิชวลไลเซชันทำให้มองเห็นด้วยภาพที่ชัดเจนยิ่งขึ้น สามารถอนุมานได้ว่า อายุและเกรดเฉลี่ยในระดับ ม.3 มีส่วนสำคัญอันดับต้นของผู้เข้ารับการศึกษา และสามารถนำไปเป็นตัวแบบในการวิเคราะห์พฤติกรรมและการจัดกลุ่มทางการเรียนต่อไปได้

คำสำคัญ: ดาต้าวิชวลไลเซชัน กราฟ การจัดกลุ่ม การทำเหมืองข้อมูล

Abstract : This study was designed based on Document Research and Explanatory Sequential Mixed Methods to represent data with visualization and to focus on selecting specific specimens. The purpose of this study was to analyze the relationship and the pattern of those who were selected as an Army Cadet from 2554-2558. The researcher proposed the concept of data analysis and classification using tree-based decision-making and statistical data visualization. There are seven main attributes for the analysis: age, GPA., degree, school, province, region and family status. These attributes have effects on being selected as a Pre-cadet. By classification by age, the highest number of students in each academic year is 17. It can be deduced that when one acquires knowledge, skill, and ability, one has high test scores and is selected. By forecasting trend data, the number of people admitted to the program was an Army Cadet. By predicting trend from linear equation, it was found that the number of selected candidates was 20.23% increase. The mean was 245 and the standard deviation was 1.79. Depending on the external factors or requests from military units which requested directly in each academic year. To classify students by the factors of grade point average of Mathayomsuksa 3, this study represents relationship between the learning outcomes by using the exponential function. When importing data using the RapidMiner data mining tool, then the results are compared. The researcher found that trends of the GPA ranging from 3.50 to 4.00, which are very good. The mean was 3.51 and the standard deviation of the data was 0.40 which was acceptable. The results of this study are part of analysis of the decision making data and the planning of the information presented to the military personnel. It can be assumed that the average age at the Mathayomsuksa 3 level is the most important and ability to learn and it can be used as model for learning group. Data visualization makes the results of analysis more justified.

Keywords: Data Visualization, Graph, Classification, Data Mining

1. บทนำ

การศึกษาทางทหารนับเป็นรากฐานเพื่อการพัฒนานายทหารที่มีความเพียบพร้อมทั้งด้านวิชาการและวิชาทหาร และในปัจจุบันทหารมีบทบาทและทำหน้าที่สำคัญอย่างยิ่ง มีความมั่นคงเป็นปึกแผ่นมาอย่างยาวนาน เริ่มต้นตั้งแต่ปีพุทธศักราช 2411 [1] สถาบันการศึกษาของทหารและตำรวจของประเทศไทยมีกระบวนการสอบและผ่านการคัดเลือกผู้สมัครที่เป็นบุคคลพลเรือนก่อนเข้าเป็นนักเรียนทหารหรือตำรวจเป็นประจำทุกปีการศึกษา ข้อความสำคัญประการหนึ่งตามคำบอกเล่าว่า “กว่าจะเป็นนักเรียนนายร้อย” ย่อมมีนัยสำคัญหลายประการ โดยคำว่า “นักเรียนนายร้อย” นั่นคือ ผู้ที่จะสำเร็จการศึกษาออกไปเป็นนายทหารสัญญาบัตรหลักของกองทัพ [2] กระบวนการหล่อหลอมจึงเน้นให้นักเรียนมีความพร้อมอย่างรอบด้าน ประมาณต้นเดือนกุมภาพันธ์ถึงเมษายนในแต่ละปีการศึกษามีจำนวนผู้สมัครสอบที่จบชั้นมัธยมศึกษาปีที่ 3 ขึ้นไป อายุไม่เกิน 17 ปี ต้องสอบผ่านการคัดเลือกเข้าเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบก โดยผ่านรอบภาควิชาการจำนวนโดยประมาณ 700 คน และผ่านการทดสอบร่างกายตรวจร่างกาย และสอบสัมภาษณ์ประมาณ 200 คน จากยอดผู้สมัครในแต่ละปีทั้งสิ้นประมาณ 20,000 คน และผนวกรวมกับกลุ่มนักเรียนนายสิบเหล่าวิทยาการต่างๆ ที่ได้รับการคัดเลือกจากผู้ที่มีความรู้ความสามารถ โดยประมาณ 20 คน ประการสำคัญคือ ผู้วิจัยได้มองย้อนกลับไปถึงภูมิหลังของนักเรียนนายร้อย ได้แก่ การศึกษาหรือการเริ่มต้นทางการศึกษา [3] คุณภาพทางด้านการเรียน ได้แก่ ผลการเรียนหรือเกรดเฉลี่ย ซึ่งข้อมูลเหล่านี้มีเรื่องราว (Story) มีความน่าสนใจอย่างยิ่งและมีความต่อเนื่องกันจนกลายเป็นข้อมูลที่มีจำนวนปริมาณมากขึ้น สามารถนำมาเล่าเรื่องราวจนเกิดการค้นพบความรู้ในฐานข้อมูล (Knowledge Discovery in Database: KDD) หรือรูปแบบข้อมูล (Pattern Data) [4] สามารถนำไปใช้ประโยชน์ในการสนับสนุนการวางแผนและตัดสินใจของผู้บังคับบัญชาได้ในอนาคต

เนื่องจากปัจจุบันได้มีการรายงานสรุปประจำปีการศึกษาและรูปแบบร้อยละ [5] เพื่อรวบรวมจัดทำสรุปเป็นลายลักษณ์อักษรหลังจากจบสิ้นกระบวนการคัดเลือกแล้ว และนำเสนอข้อมูลสารสนเทศเพื่อรายงานสรุปผลเพื่อการบริหารจัดการเบื้องต้นและผู้บังคับบัญชามองเห็นสารสนเทศในมิติต่างๆ ข้อมูล สารสนเทศ และองค์ความรู้ และสามารถทำนายแนวโน้มของการสมัครเป็นนักเรียนนายร้อยและสอบผ่านการคัดเลือกเข้ารับการศึกษา โดยข้อมูลต่างๆ เหล่านั้นถ้านำมาวิเคราะห์เชิงลึกเพื่อการค้นพบความรู้หรือพบรูปแบบใหม่ ที่อาจจะบ่งชี้ให้ทราบถึงกลุ่มของผู้สมัครสอบจึงมีความจำเป็นและเป็นประโยชน์อย่างมากต่อหน่วยงานทางทหารและตำรวจ จะเกิดประโยชน์สูงสุดเมื่อถูกนำมาวิเคราะห์ ประยุกต์ใช้ในบริบทต่างๆ โดยเฉพาะอย่างยิ่งในบริบทข้อมูลทางการศึกษา [6-8] ได้แก่ ข้อมูลของนักเรียนนายร้อยซึ่งจัดอยู่ในระดับอุดมศึกษา เป็นข้อมูลส่วนหนึ่งที่ได้รวบรวมจากการบันทึกไว้ในแต่ละปี อาจจะอยู่ในรูปแบบของไฟล์ต่างๆ ตาราง หรือฐานข้อมูล เมื่อปริมาณข้อมูลมีจำนวนเพิ่มมากขึ้นพบว่า ความรู้จากการทำเหมืองข้อมูล (Data Mining) [9] มาประยุกต์ใช้ในกระบวนการสกัดหรือสืบค้นความรู้ที่น่าสนใจในข้อมูลปริมาณมาก [10] สามารถนำมาใช้ประกอบการตัดสินใจของผู้บังคับบัญชาในการวางแผนงานทั้งระยะสั้นและระยะยาว ผู้วิจัยได้ศึกษาและประยุกต์ตัวแบบนักเรียนนายร้อยด้วยวิธีการรวมกลุ่ม (Clustering) และการค้นพบตัวแบบที่เกี่ยวข้องกันในลักษณะและสถานภาพของนักเรียนนายร้อย เพื่ออธิบายความสัมพันธ์ในองค์ประกอบของนักเรียนนายร้อย โดยจัดให้เป็นกระบวนการที่อยู่ในส่วนหลัง (Back State) และแสดงผลด้วย “กราฟดาต้าวิชวลไลเซชัน” ที่มีความสำคัญอย่างยิ่งในการเป็นตัวแทนข้อมูลสารสนเทศ [11] จัดให้เป็นกระบวนการที่อยู่ในส่วนหน้า (Front State) เพื่อถ่ายทอดการเล่าเรื่องด้วยการมองเห็นด้วยภาพได้

การศึกษาครั้งนี้มุ่งเน้นศึกษาและวิเคราะห์เอกสาร (Document Research) รายงานเป็นหลัก รายงาน

สถานภาพนักเรียนเตรียมทหาร กลุ่มตัวอย่างได้แก่ ชุดข้อมูลผู้ที่ผ่านการคัดเลือกเข้าเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบก ระหว่างปีการศึกษา 2554 – 2558 ด้วยการทำเหมืองข้อมูลและแสดงผลด้วยกราฟดาต้าวิชวลไลเซชัน เพื่อการอธิบายและการค้นพบความสัมพันธ์ของผู้ผ่านการคัดเลือกเข้าเป็นนักเรียนเตรียมทหารเท่านั้น

2. วัตถุประสงค์

เพื่อศึกษา วิเคราะห์ความสัมพันธ์ และการค้นพบรูปแบบผู้ที่ผ่านการคัดเลือกเข้าเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบก ด้วยการวิเคราะห์และจำแนกข้อมูลโดยใช้เทคนิคต้นไม้ตัดสินใจและแสดงผลด้วยกราฟดาต้าวิชวลไลเซชัน

3. เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องมีดังนี้

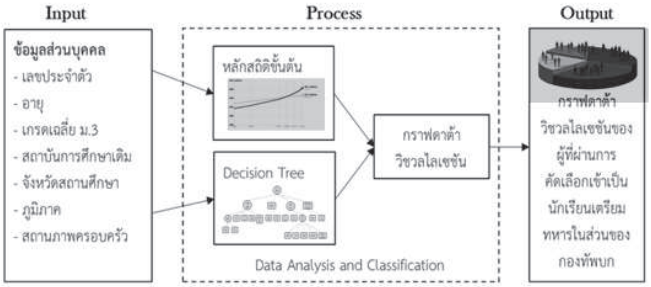
วิจักษณ์ ศรีสัจจะเลิศวาจา และดุษฎี ประเสริฐธิติพงษ์ (2557) [5] ได้นำเสนอบทบาทและความสำคัญของการทำเหมืองข้อมูลทางการศึกษา ได้อธิบายให้สามารถเข้าใจและกระชับมากขึ้นว่า การทำเหมืองข้อมูลการวิชวลไลเซชัน มีเป้าหมายสองด้าน ได้แก่ ด้านการทำนาย (Predictive Data Mining) สำหรับค้นหาตัวแบบที่อธิบายจากข้อมูลที่ได้เตรียมการไว้ ด้วยวิธีการจำแนกประเภท (Classification) ทำนาย (Prediction) หรือคาดคะเน (Estimation) ปรากฏการณ์หรือข้อมูลที่เกิดขึ้น ด้านการอธิบาย (Descriptive Data Mining) เป็นการผลิตสารสนเทศใหม่บนพื้นฐานของชุดข้อมูลที่มีอยู่ เพื่อวิเคราะห์ทำความเข้าใจข้อมูลเดิม โดยค้นหารูปแบบ และความสัมพันธ์ที่ซ่อนอยู่ภายในชุดข้อมูล จะเห็นได้ ว่าการทำเหมืองข้อมูลจะต้องอาศัยข้อมูลปริมาณมากซึ่ง อาจจะเก็บอยู่ในหลายรูปแบบ เช่น ฐานข้อมูล คลังข้อมูล ไฟล์เวิร์กชีท (Worksheet File) และเว็บ (Web) เป็นต้น ความรู้และการใช้ประโยชน์จากการทำเหมืองข้อมูลทางการศึกษานั้น ผู้บริหารสามารถนำองค์ความรู้ที่ได้มาใช้เป็นแนวทางประกอบการตัดสินใจและวางแผนงาน และให้ความเห็นว่าข้อมูลองค์ประกอบพื้นฐานการศึกษามีหลายองค์ประกอบที่มีความความสัมพันธ์ซับซ้อนและปริมาณข้อมูลมาก การทำเหมืองข้อมูลจะช่วยสกัดความรู้จากความสัมพันธ์ที่ซ่อนอยู่ในข้อมูล ค้นหากลุ่มเป้าหมายที่เข้าเรียน การทำนายสมรรถนะนักศึกษา เพื่อเป็นการแนะนำกระบวนเรียนที่เหมาะสมแก่นักศึกษา และปรับปรุงหลักสูตรให้สอดคล้อง พัฒนาศักยภาพของสถาบันการศึกษาในอนาคตได้

Matthew Michael Brehmer (2560) [6] ได้นำวิชวลไลเซชันมาใช้ในบทสรุปของงาน โดยเรียงลำดับงานในการวิเคราะห์ข้อมูล จากคำถามหลัก 2W1H ทำไมข้อมูลถึงจะมองเห็นด้วยภาพ งานอะไรบ้างที่ขึ้นต่อกันในกระบวนการอินพุตและเอาต์พุต และวิธีการแสดงงานในแง่ของการเข้ารหัสภาพและตัวเลือกการออกแบบปฏิสัมพันธ์กัน มาช่วยแก้ปัญหาการเชื่อมโยงข้อมูลความน่าสนใจ และการเข้ารหัสภาพ และตัวเลือกการออกแบบปฏิสัมพันธ์กัน การจำแนกประเภทจากเครื่องมือในการวิเคราะห์การศึกษาการสัมภาษณ์แต่ละคนที่มองเห็นมิติข้อมูลที่ลดลงด้วยในด้านต่างๆ ในโดเมนแอปพลิเคชันผลวิเคราะห์สิ่งที่พบ จากการอธิบายงานด้วยการเปรียบเทียบ และการ Cross-Pollination ตัวเลือกการออกแบบภาพในโดเมนแอปพลิเคชัน การ Mapping ใช้การประยุกต์แบบย้อนกลับ การประเมินผลการศึกษาหลังการใช้งานเครื่องมือวิเคราะห์ภาพในด้านสื่อเชิงสืบสวนและการศึกษาออกแบบวิชวลไลเซชันในด้านการจัดการพลังงานจำแนกตามลำดับงานได้ 5 หมวดหมู่ที่เกี่ยวเนื่องกันกับแสดงผลข้อมูลที่ลดลงอย่างมาก ผลคือ การจำแนกประเภทนั้นวัตถุประสงค์เพื่อแจ้งการออกแบบเครื่องมือและเทคนิคใหม่ สำหรับการแสดงผลข้อมูลนี้ ประเมินประสิทธิภาพของการจำแนกประเภทเครื่องมือวิชวลไลเซชันสำหรับการตรวจสอบการเก็บรวบรวมเอกสารข้อความขนาดใหญ่ ได้ศึกษาลักษณะงานที่เป็นนามธรรมสองแบบที่เกี่ยวกับการทำเหมืองเอกสาร และออกแบบนำเสนอเป็นบทเรียนด้วยเครื่องมือวิชวลไลเซชันสำหรับเอกสารข้อมูล

พรรณิภา บุตรเอก และสุรเดช บุญลือ (2557) [11] ได้ศึกษาตัวแบบพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาชุดข้อมูลกลุ่มตัวอย่างที่นำมาใช้ทดสอบเป็นข้อมูลของนักศึกษาระดับปริญญาตรีหลักสูตร 4 ปี สาขาวิชาวิทยาการคอมพิวเตอร์ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยนอร์ทกรุงเทพ ระหว่างปีการศึกษา 2547-2551 จำนวนทั้งสิ้น 138 ระเบียน 18 คุณลักษณะ ชุดข้อมูลที่ได้จะนำส่งเข้าตัวแบบพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา ซึ่งจะใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (SVM) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ (BP-ANN) พยากรณ์ซัพพอร์ตเวกเตอร์แมชชีนที่ใช้เคอร์เนลฟังก์ชั่นแบบ SVM-NP SVM-RBF และ SVM-PUK ให้ค่าความแม่นยำในการพยากรณ์เท่ากันที่ร้อยละ 87.68 โดยการทดสอบประสิทธิภาพตัวแบบจะใช้วิธีไขว้ทบ 15 ส่วน ส่วนตัวแบบพยากรณ์ C4.5 และ BP-ANN ให้ค่าความแม่นยำร้อยละ 83.33 และ 84.36 ตามลำดับ สรุปได้ว่าตัวแบบพยากรณ์ SVM-PK เป็นตัวแบบที่เหมาะสมที่สุดสำหรับการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา

Zhou Feiner และคณะ [12] [13] ได้แบ่งกลุ่มการนำเสนอข้อมูลออกเป็น 2 กลุ่ม ตามเป้าหมายของการใช้งาน ได้แก่ กลุ่มการแจ้งให้ทราบ ประกอบด้วย รายละเอียดของข้อมูล และการสรุปข้อมูล (Summarization) และกลุ่มการใช้งาน ประกอบด้วย การสำรวจข้อมูล (Data Exploration) เพื่อแสดงความสัมพันธ์ของข้อมูลจากแหล่งที่มา การค้นหา การจัดอันดับ วัตถุประสงค์สำคัญในการผสมผสานงานสำหรับการประเมินวิชวลไลเซชัน ได้แก่ การระบุ (Identify) โดยอ้างอิงหลักการของสถานที่ (Where) อะไรบ้าง (What) หรือ ใคร (Who) และเมื่อไร (When) [14] การเปรียบเทียบ (Compare) ระหว่างค่าหรือชนิด เวลา การจัดอันดับ (Rank) ได้แก่ อันดับหนึ่งหรืออันดับสุดท้าย ดีที่สุดหรือแย่ที่สุด ใกล้ที่สุดหรือไกลที่สุด อีกประการที่สำคัญคือ การเกี่ยวเนื่องกัน (Associate) อาทิ ความสัมพันธ์ระหว่างตัวแปรแนวโน้มกับเวลา อาทิ เป็นเหตุเป็นผลกัน

จากการทบทวนเอกสารและงานวิจัยที่เกี่ยวข้อง ทำให้ผู้วิจัยได้ออกแบบแนวคิดกระบวนการศึกษาสำหรับงานวิจัยครั้งนี้ ดังภาพที่ 1 และกำหนดตัวแปรอิสระ ได้แก่ ปัจจัยส่วนบุคคล แบ่งออกเป็น อายุ เกรดเฉลี่ย ม.3 วุฒิการศึกษาตอนสอบเข้า สถาบันการศึกษาเดิม จังหวัดสถานศึกษา สถานภาพครอบครัว และตัวแปรตาม ได้แก่ ปีการศึกษา



ภาพที่ 1 แนวคิดกระบวนการในการศึกษาวิจัยครั้งนี้

4. วิธีดำเนินการศึกษา

การศึกษาครั้งนี้ได้ออกแบบการศึกษาด้วยวิธีวิจัยแบบผสมวิธีจากเอกสาร (Document Research) และแบบการอธิบายต่อเนื่อง (Explanatory Sequential Mixed Methods) จากการศึกษาเอกสารและข้อมูล ได้แก่ วิธีการแบบเฉพาะเจาะจง เพื่อศึกษาและค้นหาข้อมูลเชิงลึกในการขยายผลการศึกษาและนำผลที่ได้มาประมวลในการวางแผนด้านการศึกษาทางทหารของผู้บังคับบัญชาสำหรับนักเรียนทหาร [15] จากการวิเคราะห์เอกสารและข้อมูล ได้ออกแบบการจัดการข้อมูลโดยร่างแนวคิดกระบวนการในการศึกษาวิจัยครั้งนี้ ประกอบด้วย กระบวนการอินพุต คือ ข้อมูลส่วนบุคคล กระบวนการวิเคราะห์ข้อมูลด้วยการจำแนกกลุ่ม ประกอบด้วยหลักสถิติขั้นต้นและแผนภาพต้นไม้ตัดสินใจ และกระบวนการเอาต์พุตด้วยการแสดงผลการมองเห็นด้วยภาพเชิงกราฟต้าวิชวลไลเซชัน

4.1 การจัดเตรียมข้อมูล

การวิจัยนี้มุ่งเน้นเพื่อศึกษาและวิเคราะห์ข้อมูลย้อนหลังจากกลุ่มข้อมูลตัวอย่างผู้ที่ผ่านการคัดเลือกเข้าเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบก ระหว่างปีการศึกษา 2554 – 2558 จำนวนข้อมูล 1,226 เรคอร์ด โดยใช้วิธีการแบบผสมผสาน ประกอบด้วย การสัมภาษณ์แบบเจาะลึกรายบุคคล (In-Depth Interview) จากผู้เชี่ยวชาญ และผู้วิจัยเป็นเครื่องมือในการศึกษาข้อมูลจากเอกสาร (Documents) แผ่นงาน (Worksheets) ฐานข้อมูล (Database) เนื้อหา (Content) แบ่งออกเป็น 3 กระบวนการ ได้แก่ การคัดเลือกข้อมูล (Select Data) การทำความสะอาดข้อมูล (Cleansing Data) และการถ่ายโอนข้อมูล (Transform Data) มีรายละเอียดดังนี้

4.1.1 การคัดเลือกข้อมูล ที่จะนำมาใช้โดยเลือกเฉพาะข้อมูลที่เกี่ยวข้อง จำนวน 7 คุณลักษณะ คือ กลุ่มตัวอย่างผู้ที่ผ่านการคัดเลือกเข้าเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบก ระหว่างปีการศึกษา 2554 – 2558 จำนวนเรคอร์ดทั้งสิ้นคือ 1,226 ได้แก่ 243 245 245 245 248 คน ตามลำดับ และกำหนดข้อมูลที่สมบูรณ์พร้อมนำไปวิเคราะห์

4.1.2 การทำความสะอาดข้อมูล เพื่อเตรียมความพร้อมข้อมูลด้วยวิธีการดังนี้

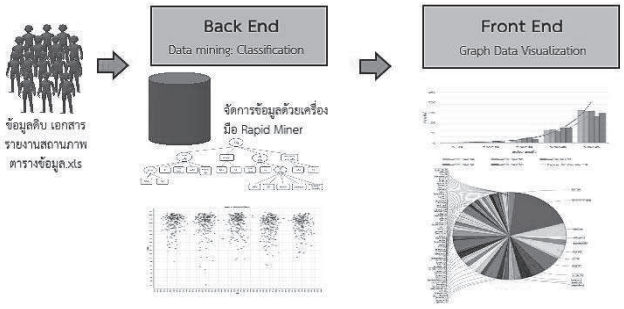
(ก) จัดการค่าว่าง (Null) โดยใส่ค่า 0

(ข) ปรับข้อมูลใส่ค่าเพื่อความเหมาะสม

(ค) จัดการค่าในคอลัมน์ “เลขประจำตัว” จะต้องเป็นค่าที่ไม่ซ้ำกัน และคอลัมน์ที่อยู่หัวแถวกำหนด “สถานศึกษาเดิม” “สถานภาพครอบครัว แทนค่าว่างด้วย อื่นๆ” ดังนั้น ทุกแถวจะต้องมีค่าเดียว

(ง) ตัดเรคอร์ดข้อมูลที่ไม่เกี่ยวข้องและไม่สมบูรณ์ออก

4.1.3 การถ่ายโอนข้อมูล นำตารางที่ใช้ในการวิเคราะห์มาเชื่อมโยงกัน ให้มีความพร้อมนำไปใช้งาน



ภาพที่ 2 แผนภาพการถ่ายโอนข้อมูล

4.2 การวิเคราะห์ข้อมูล

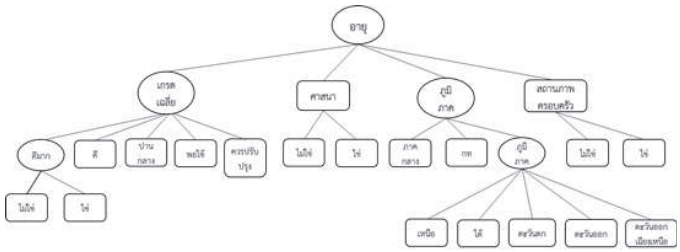
เป็นขั้นตอนเพื่อการค้นหารูปแบบในการศึกษาครั้งนี้เป็นการศึกษาแบบผสมผสานได้ออกแบบการวิเคราะห์ข้อมูล โดยใช้เทคนิคการทำเหมืองข้อมูล เพื่อแสดงความสัมพันธ์ของรูปแบบหรือความหมายที่ซ่อนเร้นอยู่ เพื่อหาข้อสรุปเชิงนามธรรมจนเกิดองค์ความรู้จากระดับล่างขึ้นสู่ระดับบน เรียกว่า “การสร้างทฤษฎีจากข้อมูล (Grounded Theory Approach)” [16] การคัดกรองที่มีความจำเป็นครั้งนี้ โดยใช้เทคนิคการจำแนกข้อมูลด้วยต้นไม้ตัดสินใจ (Decision Tree) และกราฟดาต้าวิชวลไลเซชัน เพื่อเพิ่มประสิทธิภาพในการจำแนกข้อมูลและแสดงความสัมพันธ์ของแอตทริบิวต์ นำเสนอข้อมูลด้วยกราฟแทนการใช้ข้อความ จากการกำหนดให้มีปัจจัยที่ใช้ในการทดสอบ 7 ปัจจัย ข้อมูลส่วนบุคคลที่ผ่านการจัดเตรียมและปรับปรุงแล้ว ประกอบด้วย 7 คุณลักษณะ ได้แก่ เลขประจำตัว อายุ เกรดเฉลี่ย ม.3 วุฒิการศึกษาตอนสอบเข้าสถาบันการศึกษาเดิม จังหวัดสถานศึกษา ภูมิภาค สถานภาพครอบครัว ดังตารางที่ 1

ตารางที่ 1 ปัจจัยที่ใช้นำเข้าเพื่อทำการทดสอบครั้งนี้

ลำดับ	ปัจจัย	สถานะ
1	อายุ	นำเข้า
2	เกรดเฉลี่ย	นำเข้า
3	วุฒิการศึกษาตอนสอบเข้า	นำเข้า
4	สถาบันการศึกษาเดิม	นำเข้า
5	จังหวัดสถานศึกษา	นำเข้า
6	ภูมิภาค	นำเข้า
7	สถานภาพครอบครัว	นำเข้า

หมายเหตุ: พัฒนามาจาก “ปัจจัยในการทดสอบหาความสัมพันธ์ของแอตทริบิวต์” [6]

ผู้วิจัยได้ทบทวนเอกสารงานวิจัยที่เกี่ยวข้องพบว่าได้จำแนกภูมิภาคของประเทศไทย ยึดตามอักขรานุกรมภูมิศาสตร์ ฉบับราชบัณฑิตยสถาน เล่ม 6 ภาคผนวก 2522 แบ่งออกได้เป็น 6 ภูมิภาค ได้แก่ ภาคกลาง ภาคเหนือ ภาคตะวันออกเฉียงเหนือ ภาคตะวันออก ภาคตะวันตก และภาคใต้

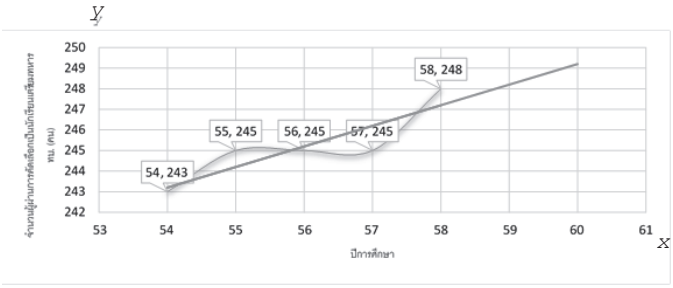


ภาพที่ 3 ออกแบบแผนภาพต้นไม้ตัดสินใจ

จากตารางที่ 1 การนำปัจจัยที่ใช้นำเข้าเพื่อทำการทดสอบ มาออกแบบการวิเคราะห์ข้อมูลและแสดงผลด้วยแผนภาพต้นไม้ตัดสินใจ ดังแสดงในภาพที่ 3

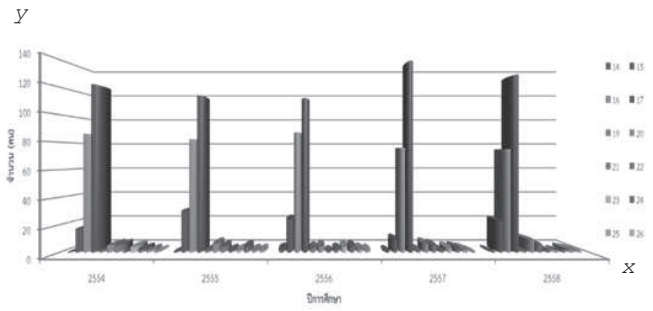
5. ผลการศึกษา

การวิเคราะห์ข้อมูลย้อนหลังของนักเรียนเตรียมทหารในส่วนของกองทัพบกที่จำแนกตามปัจจัยสามารถแสดงผลแทนค่าข้อมูลด้วยกราฟดาต้าวิชวลไลเซชัน



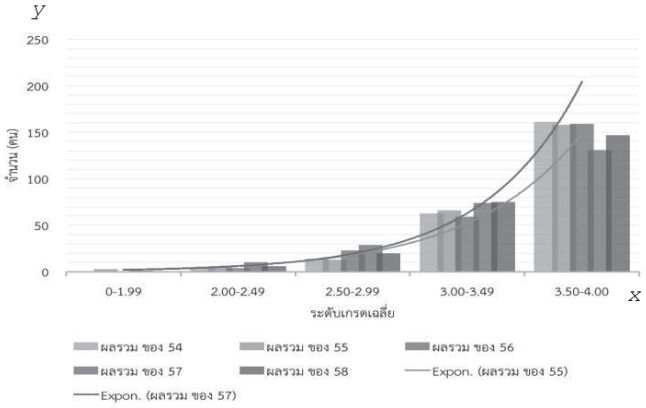
ภาพที่ 4 การพยากรณ์ข้อมูลแนวโน้มจำนวนผู้ได้เข้ารับการศึกษาเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบก ระหว่างปีการศึกษา 2554 – 2558

ค่าการพยากรณ์ข้อมูลแนวโน้มจำนวนผู้ได้เข้ารับการศึกษาเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบก ระหว่างปีการศึกษา 2554 – 2558 ด้วยวิธีคาดคะเนแนวโน้ม จากสมการแนวโน้มเชิงเส้น พบว่า ปีการศึกษา 2558 (แกน x) มีจำนวนผู้เข้ารับการศึกษาเพิ่มขึ้น จำนวน 3 คน (แกน y) คิดเป็น 20.23 เปอร์เซ็นต์ ค่าเฉลี่ยที่ได้คือ 245 และค่าเบี่ยงเบนมาตรฐานข้อมูลคือ 1.79 เนื่องจากข้อมูลมีความใกล้เคียงกัน ทั้งนี้ขึ้นอยู่กับปัจจัยภายนอกหรือความต้องการจากหน่วยขึ้นตรงร้องขอในแต่ละปีการศึกษา กราฟดาต้าวิชวลไลเซชันแสดงผลการวิเคราะห์ข้อมูลสถิติพื้นฐานจำแนกตามปัจจัยของผู้ผ่านการคัดเลือกเข้าเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบก



ภาพที่ 5 จำแนกตามอายุของผู้เข้ารับการศึกษาเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบก

จากภาพที่ 5 โดยการจำแนกตามอายุ พบว่า นักเรียนอายุ 17 ปีที่แทนข้อมูลด้วยกราฟแท่งสีม่วง คือ ผู้เข้ารับการศึกษาที่มีจำนวนสูงสุด (แกน x) ในแต่ละปีการศึกษา (แกน y) สามารถอนุมานได้ว่าเมื่อการสั่งสมประสบการณ์ความรู้ทักษะความสามารถในระดับหนึ่งแล้วจะสามารถผ่านการทดสอบได้สูง



ภาพที่ 6 จำแนกตามปัจจัยเกรดเฉลี่ยในระดับ ม.3 กับความสัมพันธ์เชิงแนวโน้มผลการเรียน

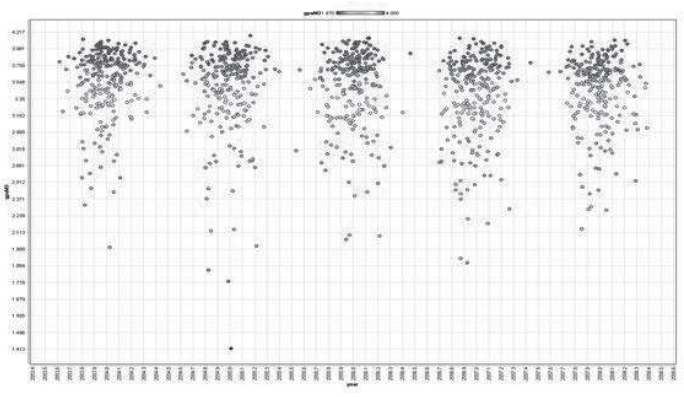
ผลการจำแนกตามปัจจัยผลการเรียนเกรดเฉลี่ยในระดับ ม.3 กับความสัมพันธ์เชิงแนวโน้มผลการเรียน ด้วยฟังก์ชันเอกซ์-โพเนนเชียล (Exponential Function) [15,16] ด้วยนิยามกราฟของฟังก์ชัน คือ

$f(x) = ax$, $a > 0$ and $a \neq 1$ (1)

ฟังก์ชัน ax เมื่อ a คือ จำนวนที่ทำให้ฟังก์ชัน ax เท่ากับอนุพันธ์ของตัวเอง (โดยที่ a มีค่าโดยประมาณเท่ากับ 2.718) มีลักษณะตั้งชันขึ้นและมีอัตราเพิ่มค่าความชันยิ่งขึ้นเมื่อ x เพิ่มขึ้น (x คือ ค่าที่เกิดขึ้นจริง, ค่าที่เกิดจากการเปลี่ยนแปลงคงตัวในตัวแปรอิสระ ส่งผลให้เปลี่ยนแปลงตามสัดส่วนทิศทางเดียวกันกับตัวแปรตาม คือ เมื่อการเพิ่มขึ้นหรือลดลงของอัตราร้อยละ) และกราฟจะวางตัวอยู่เหนือแกน x เสมอ

ในปีการศึกษา 2555 และ 2557 สามารถคาดการณ์ได้ว่า เกรดเฉลี่ยในระดับ ม.3 (แกน x) ของปีการศึกษาถัดไปของข้อมูลผู้ที่ผ่านการคัดเลือกเข้าเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบกนั้น อยู่ในระดับดีมาก คือ ช่วง 3.50-4.00 โดยในปีการศึกษา 2554 จำนวน 161 คน (แกน y) ปีการศึกษา 2555 จำนวน 158 คน ปีการศึกษา 2556 จำนวน 159 คน ปีการศึกษา 2557 จำนวน 131 คน และปีการศึกษา 2558 จำนวน 147 คน ตามลำดับ ระหว่างปีการศึกษา 2554 – 2558 พบว่า ค่าเฉลี่ยในแต่ละช่วง 0-1.99 มีค่าเฉลี่ยอยู่ที่ 1 คน ช่วง 2.00-2.49 มีค่าเฉลี่ยอยู่ที่ 6 คน ช่วง 2.50-2.99 มีค่าเฉลี่ยอยู่ที่ 20 คน ช่วง 3.00-3.49 มีค่าเฉลี่ยอยู่ที่ 67 คน และช่วง 3.50-4.00 มีค่าเฉลี่ยอยู่ที่ 151 คน และคาดคะเนได้ว่านักเรียนเตรียมทหาร คือ บุคคลที่มีความรู้ความสามารถสูงมากทั้งวิชาการและสมรรถภาพร่างกายมีความสมบูรณ์และพร้อมเข้ารับการการศึกษาทางวิชาการและทางทหาร ดังภาพที่ 6

เมื่อนำข้อมูลเข้าโดยใช้เครื่องมือการทำเหมืองข้อมูลด้วย Rapid Miner และนำผลที่ได้มาเปรียบเทียบกับการวิเคราะห์ข้อมูลด้วยหลักสถิติเบื้องต้น ด้วยการทดสอบด้วยปัจจัยเกรดเฉลี่ย ม.3 (แกน y) ในแต่ละปีการศึกษา (แกน x) พบว่า แนวโน้มของกลุ่มข้อมูลที่แสดงผลด้วยภาพปริมาณกลุ่มใกล้เคียงกันมากที่สุด จากผลการแสดงค่าข้อมูลสารสนเทศด้วยกลุ่มสีแดง คือ ช่วงระดับ 3.50-4.00 อยู่ในระดับดีมาก มีประสิทธิภาพในการจำแนกข้อมูลได้ และค่าเฉลี่ยที่ได้คือ 3.50 ค่าเบี่ยงเบนมาตรฐานข้อมูลคือ 0.40 อยู่ในเกณฑ์ที่ยอมรับได้

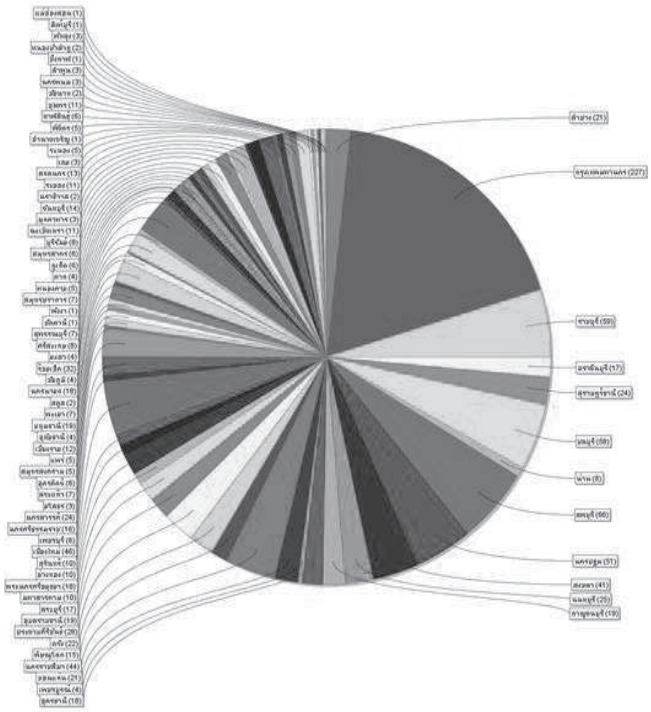


ภาพที่ 7 ผลการวิเคราะห์ข้อมูลด้วยเครื่องมือการทำเหมืองข้อมูลเพื่อจำแนกกลุ่มข้อมูลด้วยปัจจัยเกรดเฉลี่ย ม.3

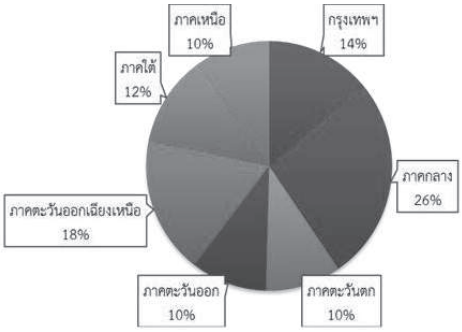
จากภาพที่ 7 จำนวนผู้ที่ผ่านการคัดเลือกเข้าเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบก ในแต่ละจังหวัดของประเทศไทย ระหว่างปีการศึกษา 2554 – 2558

จากการจำแนกด้วยจังหวัดของประเทศไทย ระหว่างปีการศึกษา 2554 – 2558 จากภาพที่ 8 นั้น พบว่ามีจำนวนผู้ที่ผ่านการคัดเลือกเข้าเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบกมากที่สุด คือ จังหวัดกรุงเทพมหานคร คิดเป็น 18.52 เปอร์เซ็นต์ รองลงมา คือ จังหวัดลพบุรี คิดเป็น 5.38 เปอร์เซ็นต์ และลำดับที่สาม คือ จังหวัดราชบุรี คิดเป็น 4.18 เปอร์เซ็นต์ จากจำนวนทั้งหมด 1,226 คน และดังภาพที่ 8 ความสัมพันธ์ระหว่างระดับวุฒิการศึกษาและภูมิภาค (แกน x) พบว่า ภาคกลางมีจำนวนผู้เข้ารับการศึกษาสูงที่สุด (แกน y) โดยใช้คุณวุฒิการศึกษาอื่นๆ สามารถอนุมานได้ว่าบุคคลผู้นั้นได้เตรียมความพร้อมทางวิชาการเพื่อสอบเข้าเป็นนักเรียนทหาร และสรุปยอดเป็นร้อยละในแต่ละภูมิภาค ระหว่างปีการศึกษา 2554 – 2558

ระหว่างปีการศึกษา 2554 – 2558 คิดเป็นร้อยละ 26 พบว่า ภูมิภาคกลางมีผู้ที่ผ่านการคัดเลือกเข้าเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบกมากที่สุดรวมกรุงเทพมหานคร รองลงมาคือ ภาคตะวันออกเฉียงเหนือ ภาคใต้ ภาคเหนือ ภาคตะวันออก และภาคตะวันตก ตามลำดับ ดังภาพที่ 9



ภาพที่ 8 ความสัมพันธ์ระหว่างระดับวุฒิการศึกษาและภูมิภาค



ภาพที่ 9 สรุปยอดเป็นร้อยละในแต่ละภูมิภาค ระหว่างปีการศึกษา 2554 – 2558

6. สรุปและอภิปรายผล

การศึกษาครั้งนี้ได้นำความรู้และการจัดการข้อมูลเทคโนโลยีสารสนเทศมาประยุกต์ใช้ในการวิเคราะห์และสังเคราะห์ข้อมูลจากเอกสารรายงานด้วยการจำแนกข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจ และแสดงผลให้มองเห็นด้วยภาพทางสายตาด้วยกราฟดาต้าวิชวลไลเซชัน พบว่า รูปแบบที่ได้สามารถนำมาสังเคราะห์ข้อมูลได้ว่า

20 อันดับแรกของผู้ที่ผ่านการคัดเลือกเข้าเป็นนักเรียนเตรียมทหารในส่วนของกองทัพบก ระหว่างปีการศึกษา 2554 – 2558 อายุเฉลี่ยแรกเข้า 14 – 15 ปี มีผลการเรียนเกรดเฉลี่ยในระดับ ม.3 ค่าเฉลี่ยที่ 3.92 ศึกษาอยู่ในภาคกลางและกรุงเทพมหานครเป็นส่วนใหญ่ และสถานภาพครอบครัวส่วนใหญ่อยู่ด้วยกัน 95 เปอร์เซ็นต์ การจำแนกตามอายุพบว่า อายุ 17 ปี คือผู้ที่เข้ารับการศึกษามีจำนวนสูงสุดในแต่ละปีการศึกษา สามารถอนุมานได้ว่า นักเรียนเตรียมทหารเป็นผู้ที่มีความรู้ ความสามารถสติปัญญา ทักษะ ความเพียรพยายามมานะและตั้งใจอย่างยิ่ง จากการสั่งสมประสบการณ์ในระดับหนึ่งแล้วจะสามารถผ่านการทดสอบทั้งทางสติปัญญา สมรรถภาพร่างกายมีประสิทธิภาพพร้อมเข้ารับการศึกษา ผลการจำแนกตามปัจจัยผลการเรียนเกรดเฉลี่ยในระดับ ม.3 กับความสัมพันธ์แนวโน้มผลการเรียนพบว่า แนวโน้มของกลุ่มข้อมูลเกรดเฉลี่ยช่วง 3.50-4.00 อยู่ในระดับดีมาก มีประสิทธิภาพในการจำแนกข้อมูล และค่าเฉลี่ยที่ได้คือ 3.51 ค่าเบี่ยงเบนมาตรฐานข้อมูลคือ 0.40 อยู่ในเกณฑ์ที่ยอมรับได้ โดยกองทัพได้บุคคลที่มีความพร้อมทั้งสติปัญญา ร่างกาย และจิตใจในการเข้ารับการศึกษาทางด้านการทหารนี้

7. กิตติกรรมประกาศ

ขอขอบคุณกองทัพบกที่ให้การสนับสนุนทุนการศึกษาเพื่อพัฒนาบุคลากรด้านการศึกษา กองสถิติและประเมินผล โรงเรียนเตรียมทหาร กองสถิติและทะเบียนประวัติ โรงเรียนนายร้อยพระจุลจอมเกล้า (กสป.รร.จปร.) ที่สนับสนุนข้อมูลเอกสารรายงาน ประกอบกับข้อคิดเห็นต่างๆ ที่เป็นประโยชน์ในการทำงานวิจัย กองวิชาวิศวกรรมไฟฟ้า ส่วนการศึกษา โรงเรียนนายร้อยพระจุลจอมเกล้า (กวฟ.สกศ.รร.จปร.) ที่ให้การสนับสนุนอำนวยความสะดวกการจัดทำเอกสาร หนังสือขอความอนุเคราะห์ และสนับสนุนเวลาในการทำงานวิจัย

8. บรรณานุกรม

(1) ประวัติโรงเรียนเตรียมทหาร. (สืบค้นจาก) http://www.afaps.ac.th. และhttps://th.wikipedia.org/wiki/โรงเรียนเตรียมทหาร.

(2) ข้อมูลโรงเรียนนายร้อยพระจุลจอมเกล้า. (สืบค้นจาก) http://www2. crma.ac.th/e_book/Thai.asp.

(3) T. Thilagaraj, N. Sengottaiyan, 2017. A Review of Educational Data Mining in Higher Education System, p. 349–358.

(4) C. C. Yang, T. Dorbin Ng, 2011. Analyzing and Visualizing Web Opinion Development and Social Interactions with Density-Based Clustering, IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, Vol. 6, p. 1144–1155.

(5) วิจักษณ์ ศรีสัจจะเลิศวาจา และ ดุษฎี ประเสริฐธิติพงษ์, 2557. การทำเหมืองข้อมูลทางการศึกษา (Educational Data Mining), วารสารวิชาการ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่, ปีที่ 20, เดือน ก.ค.-ก.ย..

(6) Matthew Michael Brehmer, 2016. Why Visualization? Task Abstraction for Analysis and Design, A Dissertation-Doctor of Philosophy, The University of British Columbia., A Dissertation-Doctor of Philosophy, The University of British Columbia.

(7) กาญจนา หฤหรรษพงศ์, 2558. แบบจำลองเพื่อช่วยทำนายผลการย้ายสาขาวิชา โดยการจำแนกประเภทข้อมูล กรณีศึกษามหาวิทยาลัยวลัยลักษณ์, The Tenth National Conference on Computing and Information Technology, NCCIT, น. 239–244.

(8) K. Kohli, S. Birla, 2016. Data Mining on Student Database to Improve Future Performance, International Journal of Computer Applications, Vol. 15, p. 42–46.

(9) ธีรพงษ์ สังข์ศรี, 2557. การวิเคราะห์พฤติกรรมสำหรับการเลือกสมัครสาขาวิชาเรียนและการเปรียบเทียบตัวแบบพยากรณ์จำนวนนักศึกษาใหม่โดยใช้เทคนิคการทำเหมืองข้อมูล, The National Conference on Computing and Information Technology, NCCIT.

(10) A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer, 2017. Vega-Lite: A Grammar of Interactive Graphics, IEEE Transactions on Visualization and Computer Graphics, Volume: 23, Issue: 1, p.341 – 350.

(11) พรรณิภา บุตรเอก และ สุรเดช บุญลือ, 2557. การพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน, Veridian E-Journal Science and Technology Silpakorn University, ฉบับที่ 6, น. 40–49.

(12) S. Henderson, S. Feiner, 2011. Exploring the benefits of augmented reality documentation for maintenance and repair, IEEE Transactions on Visualization and Computer Graphics, Vol. 10, p.1355–1368.

(13) Y. Guo, X. Zhou, A.L. Porter, D.K.R. Robinson, 2015. Tech mining to generate indicators of future national technological competitiveness: Nano-Enhanced Drug Delivery (NEDD) in the US and China, Technological Forecasting and Social Change, p.168–180.

(14) กาญจนา โชคเหรียญสุขชัย, 2560. กระบวนทัศน์การวิจัยสื่อสารมวลชน สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

(15) ภัทร์พงศ์ พงศ์ภัทรกานต์, วิชัย พัวรุ่งโรจน์, คมยุทธ ไชยวงษ์สุชาดา พรหมโคตร และปาริชาติ แสงระชัฏ, 2560. การใช้เทคนิคเหมืองข้อมูลเพื่อวิเคราะห์ปัจจัยในการใช้บริการห้องสมุดของนักศึกษา, Pulinet Journal, Vol. 4, No. 2, p.10-18.

(16) G.K. Rockswold (2012). COLLEGE ALGEBRA with Modeling & Visualization. 4 th ed. Pearson Education, Inc., p.396.