

การศึกษาเทคนิคพยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีสาขา
คอมพิวเตอร์โดยใช้เทคนิคเหมืองข้อมูล
The Study of Prediction to Career for Bachelor Graduates of
Computers Using Data Mining Techniques

วีรศักดิ์ ฟองเงิน* และสุรพงษ์ เพ็ชรหาญ

สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏรำปาง

ตำบลชมพู อำเภอเมือง จังหวัดลำปาง 52100

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อศึกษาเทคนิคพยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูลที่เหมาะสม และ 2) เพื่อเปรียบเทียบผลการวิเคราะห์พยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูล งานวิจัยนี้ได้นำข้อมูลภาวะการมีงานทำของบัณฑิต และข้อมูลประวัติของนักศึกษาระดับปริญญาตรีหลังสำเร็จการศึกษาจากสำนักงานคณะกรรมการการอุดมศึกษาย้อนหลัง 5 ปี คือปี 2555–2559 จำนวน 65,335 ระเบียบ ในสาขาวิชาทางด้านคอมพิวเตอร์ และมีคุณลักษณะประกอบด้วย ผลการเรียน ความสามารถพิเศษ อาชีพของบิดามารดา รายได้ของบิดามารดา เพศ ตำแหน่งงาน ความสอดคล้องสาขา สาขาวิชาทดลองวัดความแม่นยำด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอรัลและเทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิง ผลการวิจัยพบว่าความแม่นยำในการจำแนกประเภทข้อมูลจาก 3 เทคนิค ประกอบด้วย 1) เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจเท่ากับ 81.91% 2) เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอรัลเท่ากับ 84.29% และ 3) เทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิงเท่ากับ 81.71% พบว่าเทคนิคแรนดอมฟอรัลให้ความถูกต้องในการจำแนกประเภทข้อมูลสูงที่สุดเพราะมีค่าความถูกต้องมากที่สุด

คำสำคัญ : การแนะนำอาชีพ เหมืองข้อมูล ต้นไม้ตัดสินใจ แรนดอมฟอรัล แบ็กกิง

Abstract

The purposes of the research were 1) to study of prediction to techniques of career for undergraduate students of the computer program by using the appropriate datamining technique, and 2) to compare the results of the analysis of prediction to career for the students by using the data mining techniques. The data was collected from 65,335 employees who graduated in the computer program, Office of the Higher Education Commission during 2012 to 2016. The independent variables were grade average, talent, parental occupation, parental income, gender, job position, target and major. Decision Tree Technique, Random Forest Technique and Bagging Technique were used to analyze and classify the data. The research findings revealed that the most appropriate technique in verifying the accuracy was Random Forest Technique. The percentage of three techniques in verifying the accuracy of the data was 1) 81.91% of Decision Tree, 2) 84.29% of Random Forest and 3) 81.71% of Bagging Techniques.

Keyword : Career Guidance, Data Mining, Decision Tree, Random Forest, Bagging

บทนำ

ปัจจุบันความต้องการจ้างงานในตลาดแรงงานของประเทศไทย ขึ้นอยู่กับสภาพการเจริญเติบโตของเศรษฐกิจ ณ ช่วงเวลานั้น ๆ ทำให้การเลือกอาชีพนับว่าเป็นเรื่องสำคัญอย่างยิ่งในชีวิตมนุษย์ ต้องมีการเริ่มต้นด้วยการวางแผนตั้งแต่วัยเรียน ในสถาบันอุดมศึกษา เพื่อให้เพิ่มโอกาสในการเลือกอาชีพที่เหมาะสมกับความต้องการสอดคล้องกับความสามารถ สถาบันอุดมศึกษาจึงต้องผลิตบัณฑิตให้มีความรู้ความสามารถครบทุกด้าน พร้อมกับการส่งเสริมแนะแนวทางอาชีพให้กับนักศึกษา แต่ปัจจุบันการแนะแนวอาชีพในสถาบันอุดมศึกษาในประเทศไทย ที่ได้ดำเนินการกันอยู่ในปัจจุบันก็เป็นการให้ความช่วยเหลือเกี่ยวกับการให้คำปรึกษาทางด้านอาชีพ อาทิเช่น การเลือกอาชีพ เลือกสถานทำงาน วิธีสมัครงาน โดยนักศึกษาจะมาขอคำปรึกษาและความช่วยเหลือจากอาจารย์ที่ทำหน้าที่ให้คำปรึกษาประจำศูนย์ให้คำปรึกษา และแนะแนวของมหาวิทยาลัยและวิทยาลัยนั้น สำหรับบางมหาวิทยาลัยได้แยกบริการจัดหางานให้นิสิตนักศึกษาทำ ทั้งงานพิเศษและงานถาวรมาไว้ที่ฝ่ายกิจการนิสิตนักศึกษาของมหาวิทยาลัยและวิทยาลัย แต่อย่างไรก็ตาม การแนะแนวอาชีพในสถาบันอุดมศึกษาของประเทศไทยก็ยังไม่ได้ผลเต็มที่ (ชัชชญา 2557) นักศึกษาจำนวนมากประสบปัญหาในการตัดสินใจเลือกอาชีพให้ตรงกับตัวเอง มีเพียงจำนวนน้อยคนที่ตัดสินใจเลือกอาชีพที่ตรงกับตัวเอง การเลือกอาชีพผิดอาจจะทำให้การดำรงชีวิตยุ่งยาก (รัชดา 2552) ซึ่งอาจจะส่งผลต่อความไม่พอใจในการทำงาน ปฏิบัติหน้าที่ได้ไม่ดี เกิดความวิตกกังวล เกิดความเครียดและไม่ใส่ใจต่อสังคม (สมฤทัย มปป.) ความต้องการบุคลากรทางสายงานที่เกี่ยวข้องกับวิชาชีพคอมพิวเตอร์มีจำนวนมาก แต่ในทางกลับกันนั้นบุคลากรสายงานนี้ยังไม่เพียงพอต่อความต้องการ เนื่องจากขาดแคลนบุคลากรที่มีพื้นฐานและความรอบรู้ด้านสายงานคอมพิวเตอร์อีกจำนวนมาก ถึงแม้ว่าจะเป็นสายวิชาชีพที่นิยมเรียนในระดับอุดมศึกษาก็ตาม แต่ก็ยังไม่สามารถพัฒนาบุคลากรที่มีคุณภาพให้สอดคล้องกับความต้องการแรงงานของตลาดได้และสอดคล้องกับนโยบายประเทศที่มุ่งเน้นการพัฒนาบุคลากร ICT ทางด้านอุตสาหกรรมซอฟต์แวร์ประกอบกับเทคโนโลยีและสารสนเทศเติบโตอย่างรวดเร็วในช่วงระยะเวลาที่ผ่านและเปลี่ยนแปลงอย่างต่อเนื่อง ดังนั้นในแต่ละหน่วยงานจะต้องใช้ผู้มีความรู้และเชี่ยวชาญเฉพาะในการปฏิบัติงาน นำมาประยุกต์ใช้กับงานเฉพาะด้าน ให้สามารถใช้งานได้อย่างมีประสิทธิภาพและการเตรียมพร้อมเข้าสู่ประชาคมเศรษฐกิจอาเซียน (AEC) ที่จะสร้างมาตรฐานในการรับรองผู้ถนัดเฉพาะทาง ในตำแหน่งพิเศษด้านคอมพิวเตอร์และเทคโนโลยี เช่น นักวิเคราะห์ระบบเครือข่าย ผู้ดูแลระบบและรักษาความปลอดภัยเครือข่ายรวมถึงการออกแบบพัฒนาซอฟต์แวร์

ผู้วิจัยจึงมีความสนใจที่จะหาวิธีแก้ปัญหาดังกล่าว โดยการนำข้อมูลภาวะการมีงานทำกับข้อมูลระเบียบประวัติของผู้สำเร็จการศึกษาตั้งแต่ปี.ศ. 2555-2559 จำนวน 65,335

ระเบียบในสาขาวิชาทางด้านคอมพิวเตอร์ ของมหาวิทยาลัยในประเทศไทยมาศึกษาเทคนิคการพยากรณ์อาชีพโดยใช้เทคนิคเหมืองข้อมูล (Data Mining) ในการจำแนกข้อมูล (Classification) เพื่อเป็นการเพิ่มความแม่นยำในการจำแนกข้อมูลให้มีความถูกต้องมากที่สุด จึงจำเป็นต้องเลือกใช้เทคนิควิธีจำแนกข้อมูลที่เหมาะสมกับข้อมูลและได้ค่าความแม่นยำที่อยู่ในระดับที่ยอมรับได้ ด้วยการใช้คุณลักษณะทางด้านความสามารถหรือผลการเรียนของผู้สำเร็จการศึกษาที่ได้รับจากหลักสูตรด้านความสามารถพิเศษ ด้านเพศ ตำแหน่งงานด้านความสอดคล้องสาขา ด้านสาขาวิชา งานวิจัยนี้ได้เสนอเทคนิคการจำแนกข้อมูล 3 เทคนิค ได้แก่ เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอรัลเรส (Random Forest) และเทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิง (Bagging) เนื่องจากเป็นเทคนิคที่นิยมนำมาใช้ในงานวิจัยที่เกี่ยวข้องมากและผลความแม่นยำอยู่ในระดับดี สำหรับการวิเคราะห์ค่าความแม่นยำในการพยากรณ์ และนำเทคนิคที่ให้ค่าความแม่นยำมากที่สุดไปใช้เป็นตัวแบบในการพัฒนาระบบพยากรณ์เพื่อการแนะแนวอาชีพต่อไป

วัตถุประสงค์การวิจัย

- 1) เพื่อศึกษาเทคนิคพยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูล
- 2) เพื่อเปรียบเทียบผลการวิเคราะห์พยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูล

เอกสารและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

ความหมายของการแนะแนวอาชีพ

การแนะแนวอาชีพ (สมฤทัย มปป.) เป็นการช่วยเหลือบุคคลให้สามารถเลือกอาชีพที่เหมาะสมกับตนเองมาเป็นการแนะแนว เพื่อให้บุคคลมีทักษะในด้านต่าง ๆ ที่เกี่ยวกับการเลือกอาชีพและการเตรียมตัวประกอบอาชีพ ตลอดจนการพัฒนาตนเองให้มีความเจริญในการประกอบอาชีพ ทักษะที่เกี่ยวกับอาชีพดังกล่าว ได้แก่ ทักษะในการวิเคราะห์ตนเอง ทักษะในการแสวงหาและการวิเคราะห์ข้อมูลด้านอาชีพ ทักษะในการเลือกการตัดสินใจ และเตรียมตัวประกอบอาชีพ

การแนะแนวอาชีพ (รัชดา 2552) เป็นกระบวนการช่วยเหลือบุคคลและสิ่งแวดล้อม เพื่อให้สามารถนำตนเองได้ เป็นต้นว่า การตัดสินใจว่าจะศึกษาด้านใด

ประกอบอาชีพอะไร หรือแก้ปัญหาอย่างไรเมื่อเจอกับสถานการณ์ต่าง ๆ รู้จักข้อมูลทางอาชีพ สามารถตัดสินใจเลือกอาชีพได้อย่างเหมาะสม รู้จักวางแผนในการเตรียมตัวเข้าสู่โลกของอาชีพ เข้ากับอาชีพของตน สามารถปรับตัวได้อย่างมีความสุข สามารถพัฒนาตนเองในทุก ๆ ด้าน ได้อย่างถูกต้องและเหมาะสม เป็นประโยชน์ต่อส่วนรวม และสังคมได้อย่างเต็มที่ เพื่อให้มีความสำเร็จในอาชีพที่ตนเลือกซึ่งจะส่งผลถึงการพัฒนาตนเองและสังคมไปจนถึงประเทศชาติอีกด้วย ดังนั้นการแนะนำอาชีพจำเป็นสำหรับนักแนะแนว

จากความหมายของการแนะนำอาชีพที่กล่าวมาข้างต้น กล่าวได้ว่าการแนะนำอาชีพเป็นกระบวนการสำคัญ ในการส่งเสริม แนะนำ แนะนำ ให้คำปรึกษา เสนอข้อมูล ข่าวสาร สถิติ ให้บุคคลเลือกอาชีพให้ตรงกับบุคลิกภาพสอดคล้องกับความต้องการ ความรู้ ความสามารถ เพื่อให้ประสบความสำเร็จในการดำเนินชีวิต และการประกอบอาชีพที่เลือกนั้นสามารถทำประโยชน์ให้แก่ตนเองและประเทศชาติ

เหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล (เอกลิทธิ 2557) เป็นกระบวนการค้นหาความสัมพันธ์รูปแบบและแนวโน้มใหม่ ๆ โดยการใช้วิธีทางคณิตศาสตร์และสถิติวิเคราะห์ข้อมูล กระบวนการ ขั้นตอน หรือวิธีการ เทคนิครวมไปถึงซอฟต์แวร์ที่ใช้ในการวิเคราะห์ข้อมูล เพื่อค้นหาคำตอบ หรือองค์ความรู้ เพื่อนำมาใช้ประโยชน์ในการตัดสินใจ ทำนาย คาดการณ์ หรือวางแผน ในเรื่องต่าง ๆ ถือเป็นเครื่องมือที่สร้างคุณค่าให้แก่ข้อมูล นำมาซึ่งองค์ความรู้อันเป็นคุณประโยชน์ต่อหน่วยงาน องค์กร ในการบริหารจัดการเสริมสร้างความแข็งแกร่ง ให้แก่องค์กร ส่งผลให้ได้รับความนิยม และถูกนำไปประยุกต์ใช้งานอย่างแพร่หลายในปัจจุบันนั่นเอง นอกจากนี้ เหมืองข้อมูล ยังหมายรวมถึง ชุดซอฟต์แวร์ที่ใช้ในการวิเคราะห์ข้อมูล ซึ่งถูกออกแบบมาเพื่อการประยุกต์กับความต้องการใช้งานสารสนเทศในระดับสูง เช่น ระบบสนับสนุนการตัดสินใจ (Decision Support System: DSS) หรือระบบสารสนเทศสำหรับผู้บริหาร (Executive Information System: EIS) เพื่อให้ได้ซอฟต์แวร์ที่มีความสมบูรณ์พร้อมทั้งในเรื่องของการค้นหา การจัดทำรายงาน และสนับสนุนการจัดการด้านต่าง ๆ จัดเป็นเครื่องมือชิ้นใหม่ที่สามารถค้นหาข้อมูลในฐานข้อมูลขนาดใหญ่หรือข้อมูลที่เป็นประโยชน์

การทำเหมืองข้อมูล (สุรพงศ์ 2559) คือกระบวนการค้นหาสารสนเทศ หรือข้อความรู้ที่มีอยู่ในฐานข้อมูลขนาดใหญ่ที่ซับซ้อน เพื่อนำข้อความรู้ที่ได้ไปใช้ประโยชน์ในการตัดสินใจ สารสนเทศที่ได้อาจนำมาสร้างการพยากรณ์ หรือสร้างตัวแบบสำหรับการจำแนกหน่วยหรือกลุ่ม หรือแสดงความสัมพันธ์ระหว่างหน่วยต่างๆหรือให้ข้อสรุปของสาระในฐานข้อมูล การทำเหมืองข้อมูลประกอบขึ้นด้วยการนำกระบวนการทางสถิติและการเรียนรู้ผ่านระบบคอมพิวเตอร์ เพื่อสร้างตัวแบบ กฎเกณฑ์ รูปแบบ การพยากรณ์และข้อความรู้จากฐานข้อมูลขนาดใหญ่ โดยการทำเหมืองข้อมูลมีขั้นตอนการดำเนินงานหลาย

ขั้นตอนซึ่งต้องอาศัยเทคนิคหรือวิธีการต่าง ๆ เช่น วิธีการจัดกลุ่ม การค้นหาความสัมพันธ์ การพยากรณ์ เป็นต้น

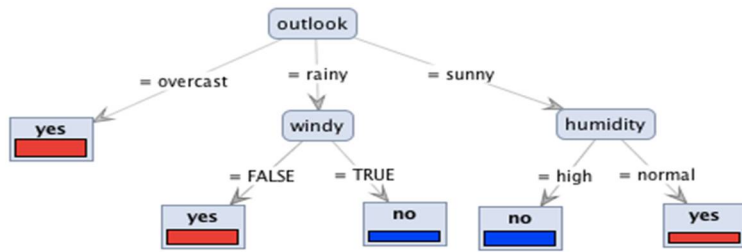
การทำเหมืองข้อมูล (สายชล 2558) คือ Application สำหรับการวิเคราะห์ประมวลผลข้อมูลที่ถูกออกแบบมาเพื่อใช้งานในระบบสนับสนุนการตัดสินใจของผู้ใช้ เป็น software ที่มีความสมบูรณ์ทั้งในเรื่องการค้นหา การสร้างรายงาน และการจัดการ ซึ่งคล้ายกับระบบ Executive Information System (EIS) หรือระบบข้อมูลสำหรับการตัดสินใจของผู้บริหาร ซึ่งเป็นเครื่องมือหนึ่งที่สามารถค้นหาข้อมูลในฐานข้อมูลขนาดใหญ่ หรือข้อมูลที่เป็นประโยชน์ในการบริหาร เป็นเพิ่มคุณค่าให้กับฐานข้อมูลที่มีอยู่

จากความหมายของเหมืองข้อมูล สรุปได้ว่า เหมืองข้อมูล หรือการทำเหมืองข้อมูล (Datamining) คือ กระบวนการสกัด หรือค้นหาสารสนเทศ เพื่อให้ได้ความรู้ หรือสารสนเทศบางมุมที่ซ่อนเร้นอยู่ในฐานข้อมูลขนาดใหญ่โดยอาศัยข้อมูลในอดีตเป็นจำนวนมาก เพื่อนำความรู้ที่ได้ไปช่วยในการวิเคราะห์ และประกอบการตัดสินใจของผู้บริหารในธุรกิจด้านต่าง ๆ ในการบริหารซึ่งเป็นการเพิ่มคุณค่าให้กับฐานข้อมูลที่มีอยู่

เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree)

การสร้างโมเดลด้วยต้นไม้ตัดสินใจ (สายชล 2558) ต้นไม้ตัดสินใจ (Decision Tree) นั้นมีความจำเป็นในการทำงานวิจัยนี้ เนื่องจากต้นไม้การตัดสินใจนั้นมีส่วนช่วยให้การวิเคราะห์ข้อมูลที่มีจำนวนมาก ๆ ได้ โดยจำลองความเป็นไปได้ของสิ่งที่เราสนใจและเงื่อนไขทั้งหมดที่ได้จากข้อมูลแต่ละตัวทำให้มองเห็นภาพรวมและวิธีการที่ต้องการได้ดียิ่งขึ้น และต้นไม้การตัดสินใจยังช่วยในการตัดสินใจทำสิ่งต่าง ๆ โดยใช้วิธีการที่ดีที่สุดได้ง่ายขึ้นกว่าการไม่ได้ใช้ต้นไม้การตัดสินใจขั้นตอนวิธีการสร้างต้นไม้ตัดสินใจ (Decision Tree Algorithm) เป็นหนึ่งในเทคนิคการทำเหมืองข้อมูลเพื่อการแบ่งประเภทข้อมูล โดยขั้นตอนวิธีนี้ จะทำการสร้างตัวแบบการทำนายที่มีลักษณะเป็นโครงสร้างต้นไม้ซึ่งประกอบไปด้วยโหนดราก (Root Node) โหนดภายใน (Internal Node) และโหนดใบ (Leaf Node)

ต้นไม้ตัดสินใจ (Ansari, 2018) เป็นวิธีการที่ได้รับความนิยมอย่างแพร่หลายเนื่องจากโมเดลที่ได้สามารถแปลความหมายและเข้าใจง่าย ลักษณะของรูปแบบข้อมูล (Pattern) ซึ่งแต่ละโหนด (Node) จะแสดงคุณลักษณะหรือแอตทริบิวต์ (Attribute) ที่ใช้ทดสอบข้อมูล แต่ละกิ่งจะแสดงผลในการทดสอบ และลีฟโหนด (Leaf Node) จะแสดงกลุ่มหรือคลาส (Class) ที่กำหนดไว้ (ชัชชญา 2557)



ภาพที่ 1 ตัวอย่างต้นไม้ตัดสินใจ (Decision Tree)

(ที่มา : จากโปรแกรม Rapid Miner Studio)

เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอว์เรส (Random Forest)

การจำแนกข้อมูลด้วยวิธีแรนดอมฟอว์เรสเป็นชุดของการจำแนกประเภทแบบไม่ตัดแต่งกิ่ง (unpruned) หรือต้นไม้ถดถอย (Regression Trees) ซึ่งถูกสร้างจากการนำข้อมูลฝึกสอนไป สุ่มเลือกตัวอย่างข้อมูลและคุณลักษณะข้อมูลแล้วนำมาสร้างเป็นต้นไม้ตัดสินใจซึ่งมีตัวอย่างส่วนหนึ่งที่ไม่ถูกเลือกเรียกข้อมูลส่วนนี้ว่า Out-of-Bag (OOB) จะถูกนำมาใช้ในการทดสอบต้นไม้ตัดสินใจ (เอกสิทธิ์ 2560) และนำผลการทำนายที่ได้มาโหวต (Vote) เพื่อหาคำตอบสุดท้าย (Ansari, 2018)

เทคนิควิธีตัวจำแนกป่าแบบสุ่ม (Random Forest) เปรียบเทียบกับวิธีการคัดเลือกตัวแปรดั้งเดิมที่ใช้ดัชนีที่ไม่ได้นำความสัมพันธ์ระหว่างตัวแปรเข้ามามี ส่วนในการคัดเลือกตัวแปร ได้แก่ สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Correlation Coefficient) การทดสอบของเวลช์ (Welch Test) เกณฑ์ของโกลับ (Golub Criterion) ผลที่ได้พบว่าวิธีการใหม่ที่น่าเสนอให้ผลที่ดีกว่าวิธี การคัดเลือกตัวแปรแบบดั้งเดิม โดยรูปแบบที่ให้ผลที่ดีคือการร่วมกันระหว่างการทำหนดจำนวนองค์ประกอบตามเกณฑ์ค่า p-value และการเลือกเซตย่อยโดยใช้กระบวนการค้นหาแบบป็นเขา และเมื่อใช้องค์ประกอบจากแนวทางดังกล่าวจะให้ผลดีกว่าเมื่อใช้ตัวแปรโดยตรงเล็กน้อย Slim PLS ในรูปแบบการร่วมกันดังกล่าวให้ผลดีกว่าการคัดเลือกตัวแปรแบบดั้งเดิมอย่างชัดเจนเมื่อจำแนกประเภทโดยใช้วิธี KNN (เอกสิทธิ์ 2557)

เทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิง (Bagging)

การจำแนกข้อมูลด้วยวิธีแบ็กกิงเป็นวิธีการที่สร้างตัวแบบแต่ละตัวที่ใช้ชุดข้อมูลเรียนรู้ที่สุ่มมาจากชุดข้อมูลเรียนรู้ที่กำหนดให้ในแบบคืน (Sampling with Replacement) กล่าวคือเมื่อสุ่มข้อมูลขึ้นมาได้หนึ่งตัวจะคืนข้อมูลนั้นกลับเข้าไปยังชุดข้อมูลเดิม ทำให้ข้อมูลนั้นมีโอกาสถูกสุ่มซ้ำอีกในอนาคต ข้อมูลแต่ละตัวในชุดข้อมูลเรียนรู้จะมีโอกาสถูกสุ่มเพื่อใช้ในการสร้างตัวแบบเท่ากับ $1 - (1 - \frac{1}{n})^n$ โดยที่ n เป็นจำนวนข้อมูลในชุดข้อมูลเรียนรู้ที่กำหนดให้ ถ้าทำการสุ่มข้อมูลด้วยวิธีนี้ n ครั้งเพื่อสร้างชุดข้อมูลสำหรับสร้างตัวแบบ 1 ตัวจะได้ชุดข้อมูลที่เป็น subset ของชุดข้อมูลเรียนรู้ที่กำหนดให้

ทั้งนี้เนื่องจากข้อมูลเดียวกันอาจถูกสุ่มซ้ำ ทำให้ชุดข้อมูลจากการสุ่มมีจำนวนข้อมูลไม่ถึง n ตัว โดยเฉลี่ยแล้วจะมีขนาดประมาณ 63.2% ของ n (Gorad et al., 2017)

จากเอกสารที่กล่าวมาสรุปได้ว่าเหมือนข้อมูลคือกระบวนการค้นหาสิ่งที่มีประโยชน์ที่สนใจ ที่เป็นองค์ความรู้จากฐานข้อมูลขนาดใหญ่ เพื่อใช้ในการวิเคราะห์หรือทำนายสิ่งต่าง ๆ ที่จะเกิดขึ้น ซึ่งเทคนิคที่ใช้ในงานวิจัยมีดังนี้ เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอเรส (Random Forest) และเทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิง (Bagging) เพื่อวัดประสิทธิภาพของโมเดล

2.2 งานวิจัยที่เกี่ยวข้อง

ชัชชญา วันดี (2557) (อ้างอิงจาก Ross 1993) ศึกษาวิจัยเรื่อง เปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลการเลือกอาชีพของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษา โดยในงานวิจัยนี้ได้ใช้ชุดข้อมูลภาวะการมีงานทำของบัณฑิต และข้อมูลระเบียบประวัติของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษา คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม ระหว่างปี พ.ศ. 2550-2554 จำนวน 12 คุณลักษณะ และ 2,515 ระเบียบ ซึ่งได้นำเทคนิคแบบจำลองต้นไม้ตัดสินใจ (Decision Tree) เทคนิคโครงข่ายประสาทเทียม (Neural Network) และเทคนิคการเรียนรู้แบบเบย์ (Naïve Bayes) มาทำการเปรียบเทียบประสิทธิภาพ ผลจากการศึกษาพบว่าประสิทธิภาพในการจำแนกข้อมูลแบบต้นไม้ตัดสินใจ มีประสิทธิภาพในการจำแนกสูงสุดด้วยค่าเฉลี่ย 88.62% และปัจจัยสำคัญที่ทำให้การเลือกอาชีพตรงหรือไม่ตรงกับสาขา มี 4 ปัจจัย คือสาขาวิชาที่เรียน เกรดเฉลี่ยเฉพาะวิชาสาขา เพศ และเกรดเฉลี่ยรวม

สมฤทัย กลัดแก้ว (มปป.) และ Breiman et al. (2001) การศึกษามีวัตถุประสงค์ 1) เพื่อศึกษาปัจจัยที่มีผลต่อการเลือกตำแหน่งงานให้สอดคล้องกับความสามารถของบัณฑิตด้วยการวิเคราะห์การถดถอยโลจิสติกพหุกลุ่มและการจำแนกประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจ 2) เพื่อเปรียบเทียบความถูกต้องของการพยากรณ์ระหว่างการวิเคราะห์การถดถอยโลจิสติกพหุกลุ่ม (Logistic Regression) และการจำแนกประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจ 3) เพื่อพัฒนาตัวแบบการตัดสินใจการเลือกตำแหน่งงานให้สอดคล้องกับความสามารถของบัณฑิต กลุ่มตัวอย่างที่ใช้ในการศึกษา ได้แก่ ข้อมูลภาวะบัณฑิตที่มีงานทำที่เข้ารับพระราชทานปริญญาบัตรในปีการศึกษา 2555-2557 จำนวน 1,933 คน เครื่องมือที่ใช้ในการศึกษา ได้แก่ แบบสำรวจข้อมูลภาวะบัณฑิตมีงานทำของศูนย์คอมพิวเตอร์มหาวิทยาลัยราชภัฏหมู่บ้านจอมบึง สถิติที่ใช้ในการวิเคราะห์ข้อมูล ได้แก่ การวิเคราะห์การถดถอยโลจิสติกพหุกลุ่มและการจำแนกประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจ พบว่าเทคนิคการจำแนกประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจมีค่าความถูกต้องมากกว่าการวิเคราะห์การถดถอยโลจิสติกพหุกลุ่มเล็กน้อย โดยค่าความถูกต้องของเทคนิคการจำแนก

ประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจเท่ากับ 57.37% และการวิเคราะห์การถดถอยโลจิสติก ทุกกลุ่มมีค่าความถูกต้อง 56.3%

Sathyavathi et al. (2016) ศึกษาวิจัยเรื่อง เหมืองข้อมูลสำหรับสร้างตัวแบบ ข้อมูลเพื่อการตัดสินใจสำหรับการพยากรณ์อาชีพ ผลการวิจัยพบว่าปัจจัยที่มีผลต่อการตัดสินใจในการเลือกอาชีพโดยการใช้เครื่องมือเหมืองข้อมูลแบบจำแนกกลุ่ม โดยใช้ขั้นตอน ในการดำเนินงานการทำเหมืองข้อมูลแบบจำแนกกลุ่ม ใช้โปรแกรม Rapidminer Studio 7 เป็นเครื่องมือ ขั้นตอนวิธีที่ใช้สำหรับการเปรียบเทียบในการจัดกลุ่มข้อมูลที่มี ประสิทธิภาพที่ได้ค่าสูงสุดเมื่อเปรียบเทียบจากขั้นตอนวิธีดังกล่าวจะพบว่า Decision tree ได้ให้ค่าความถูกต้องสูงที่สุดในการพยากรณ์สำหรับการจำแนกกลุ่มข้อมูล

Campagni et al. (2015) ศึกษาวิจัยเรื่อง ตัวแบบการทำเหมืองสำหรับอาชีพ นักศึกษา ผลการวิจัยพบว่าระเบียบวิธีการทำเหมืองข้อมูลเพื่อวิเคราะห์อาชีพของนักศึกษา ในมหาวิทยาลัย ความแตกต่างทางวิธีการจัดกลุ่มและลำดับรูปแบบของเทคนิคการจำแนก กลวิธีในการปรับปรุงประสิทธิภาพของนักศึกษาตามตารางการสอบ โดยจะแนะนำถึงอาชีพ ในอุดมคติก่อน นักศึกษาจะต้องทำการทดสอบหลังจากเรียนจบให้สอดคล้องกับหลักสูตรที่ เรียนโดยทันที แล้วเปรียบเทียบกับนักศึกษาทั่วไปกับนักศึกษาที่ใช้เทคนิคว่ามีความ แตกต่างกันและประยุกต์ใช้วิธีการดังกล่าวกับนักศึกษาจริง พบว่านักศึกษาที่ได้รับแนวคิด อาชีพในอุดมคติจะมีผลการเรียนในเทอมและจบการศึกษาที่ดีขึ้น

จากงานวิจัยที่เกี่ยวข้องผู้วิจัยได้เลือกใช้ตัวแบบเทคนิควิธีที่มีประสิทธิภาพในการ จำแนกประกอบด้วยวิธีเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจเทคนิควิธีแรนดอม ฟอรัลเรสและเทคนิควิธีแบ็กกิงเพื่อวัดประสิทธิภาพของโมเดล และเลือกข้อมูลที่น่าสนใจ ในการวิเคราะห์ซึ่งเป็นปัจจัยที่เกี่ยวข้องประกอบด้วยระเบียบประวัติผลการเรียนของนักศึกษา ประกอบไปด้วยคุณลักษณะ เพศ อาชีพบิดา รายได้บิดา อาชีพมารดา รายได้มารดา ผล การเรียนเฉลี่ย สาขาวิชา ความสามารถพิเศษ ความสอดคล้องของอาชีพกับสาขาวิชา และ ประเภทตำแหน่งงาน

วิธีดำเนินการวิจัย

1. ขั้นตอนการดำเนินการวิจัยประกอบด้วย

1. ศึกษาเทคนิคการพยากรณ์อาชีพที่เป็นไปได้และเหมาะสมในการนำมาพยากรณ์ อาชีพจากเอกสารและงานวิจัยที่เกี่ยวข้องและสรุปผลเลือกเทคนิคที่เหมาะสม

2. จัดเตรียมข้อมูล โดยประกอบด้วยข้อมูลภาวะการมีงานทำกับข้อมูลระเบียบ ประวัติของผู้สำเร็จการศึกษาตั้งแต่ปี พ.ศ. 2555-2559 จำนวน 5,335 ระเบียบ ใน สาขาวิชาทางด้านคอมพิวเตอร์ ของมหาวิทยาลัยในประเทศไทยเพื่อนำมาวิเคราะห์ด้วย เทคนิคที่เลือกไว้ในขั้นตอนที่ 1 โดยมีขั้นตอนตามมาตรฐาน 6 ขั้นตอนดังนี้

1) ทำความเข้าใจเกี่ยวกับภาวะการมีงานทำของบัณฑิต โดยศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง เทคนิคการพยากรณ์ความสัมพันธ์ข้อมูลการแนะแนวอาชีพสำหรับนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูล

2) ทำความเข้าใจข้อมูล ข้อมูลที่นำมาใช้ในการวิจัยในครั้งนี้ เป็นข้อมูลภาวะการมีงานทำของบัณฑิต และระเบียบประวัติผลการเรียนของนักศึกษา ประกอบไปด้วยคุณลักษณะ เพศ อาชีพบิดา รายได้บิดา อาชีพมารดา รายได้มารดา ผลการเรียนเฉลี่ยสาขาวิชา ความสามารถพิเศษ ความสอดคล้องของอาชีพกับสาขาวิชา และประเภทตำแหน่งงานแสดงดังตารางที่ 1

ตารางที่ 1 ตารางรายละเอียดคุณลักษณะนำมาสร้างตัวแบบ

ที่	ชื่อคุณลักษณะ	คำอธิบาย	ที่	ชื่อคุณลักษณะ	คำอธิบาย
1	GENDER_ID	เพศ	6	GPA	ผลการเรียนเฉลี่ย
2	FAT_OCCUP_ID	อาชีพบิดา	7	PROGRAM_ID	สาขาวิชา
3	FAT_REVENUE_ID	รายได้บิดา	8	QN_TALENT	ความสามารถพิเศษ
4	MOT_OCCUP_ID	อาชีพมารดา	9	TARGET	ความสอดคล้องสาขาวิชา
5	MOT_REVENUE_ID	รายได้มารดา	10	QN_POS_ID	ประเภทตำแหน่งงาน(Class)

3) การเตรียมข้อมูล

3.1) คัดเลือกข้อมูล (Data Preprocessing) นำข้อมูลจากแต่ละตารางมาทำการรวมกันโดยข้อมูลจะประกอบด้วยข้อมูลระเบียบประวัติและข้อมูลภาวะการมีงานทำ

3.2) จัดรูปแบบข้อมูล (Data Transformation) ทำการจัดรูปแบบข้อมูลเพื่อให้ข้อมูลมีความพร้อมที่จะนำไปใช้งานต่อไป

3.3) การแบ่งส่วนข้อมูล การแบ่งข้อมูลออกเป็นส่วนหลายส่วนเท่าๆ กันแล้วนำข้อมูลที่แบ่งไว้ไปใช้ในการสร้างตัวแบบและการทดสอบตัวแบบ เวียนกันไปจนครบทุกส่วนของข้อมูล การกำหนดจำนวนส่วนของข้อมูลแบ่งออกเป็น 10 ส่วน (10 folds Cross-Validation) คือการแบ่งข้อมูลออกเป็น 10 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดล ทำวนไปเช่นนี้จนครบจำนวนที่แบ่งไว้โดยวิธีนี้เป็นทดสอบประสิทธิภาพของโมเดลเนื่องจากผลที่ได้มีความน่าเชื่อถือในการวัดประสิทธิภาพ

4) การวิเคราะห์ข้อมูลและสร้างตัวแบบโดยการใช้เทคนิคที่ศึกษา

5) การแปลผลและประเมินผล ทำการประเมินผลตัวแบบจากแต่ละเทคนิค

6) การนำตัวแบบไปประยุกต์ใช้

2. เครื่องมือการวิจัย

1. โปรแกรมที่ใช้ในการวิเคราะห์ข้อมูลการพยากรณ์อาชีพได้แก่โปรแกรม WEKA
2. เทคนิคการพยากรณ์ จำนวน 3 เทคนิค ได้แก่ เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอรัล (Random Forest) และเทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิง (Bagging)

3. สถิติที่ใช้ในการวิจัย

เป็นสถิติเพื่อหาค่าความถูกต้อง และ หาค่าความคลาดเคลื่อนการทำนายข้อมูลประกอบด้วย

1. สถิติแบบความแตกต่างของค่ากลางของสองประชากรไม่อิสระ (Paired T-Test)

2. สถิติการพยากรณ์โดยพิจารณาค่าความถูกต้องที่มีค่าสูงที่สุด (Accuracy)

$$\text{Accuracy} = 100 - \% \text{Error} \quad (1)$$

3. ความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (2)$$

4. ค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ (MAE) ที่มีค่าต่ำที่สุด

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (3)$$

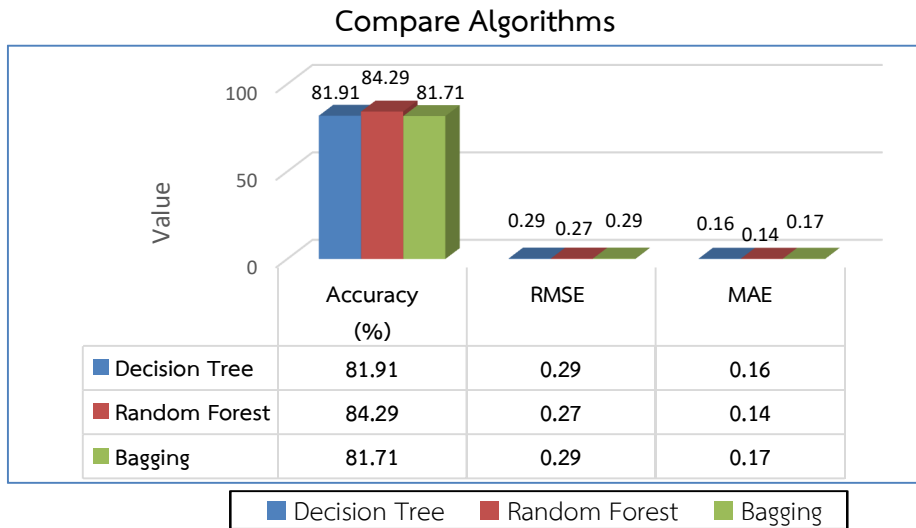
ผลการวิจัย

1. ผลการศึกษาเทคนิคการพยากรณ์จำแนกข้อมูลโดยใช้เทคนิคเหมืองข้อมูล

ผู้วิจัยได้ศึกษาเทคนิคเหมืองข้อมูลการจำแนกข้อมูล โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอรัล เทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิง และเทคนิคการจำแนกข้อมูลด้วยวิธีบูสต์ติงวิเคราะห์เปรียบเทียบกับข้อมูลภาวะการมีงานทำใช้ค่าความถูกต้องความคลาดเคลื่อนเฉลี่ยกำลังสอง ค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ เพื่อวัดประสิทธิภาพของตัวแบบ ได้ผลการดำเนินงานดังตารางที่ 2 และภาพที่ 2

ตารางที่ 2 ตารางการเปรียบเทียบผลการวิเคราะห์ค่าความถูกต้องของแต่ละเทคนิค

ที่	Algorithms	Accuracy (%)	RMSE	MAE
1	Decision Tree	81.91	0.29	0.16
2	Random Forest	84.29	0.27	0.14
3	Bagging	81.71	0.29	0.17



ภาพที่ 2 แสดงกราฟการเปรียบเทียบผลการวิเคราะห์ค่าความถูกต้องของแต่ละเทคนิค

จากตารางที่ 2 พบว่า ตัวแบบการพยากรณ์อาชีพด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอเรสให้ค่าความถูกต้องสูงที่สุดที่ร้อยละ 84.29 ตามด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจให้ค่าความถูกต้องร้อยละ 81.91 และเทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิงให้ค่าความถูกต้องร้อยละ 81.71 มีความคลาดเคลื่อนเฉลี่ยกำลังสองเท่ากับ 0.27, 0.29 และ 0.29 มีค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์เท่ากับ 0.14, 0.16 และ 0.17 ตามลำดับ เปรียบเทียบค่าความถูกต้องและนำผลไปใช้ในการดำเนินการในขั้นตอนต่อไป

2. ผลการเปรียบเทียบเทคนิคการพยากรณ์โดยใช้เทคนิคเหมืองข้อมูล

จากการเปรียบเทียบผลการวิเคราะห์ค่าความถูกต้องของแต่ละเทคนิคสร้างตัวแบบการพยากรณ์อาชีพจากข้อมูลภาวะการมีงานทำของบัณฑิตนำมาทดสอบค่าทางสถิติแบบความแตกต่างของค่ากลางของสองประชากรไม่อิสระ (Paired T-Test) พบว่าเทคนิคการจำแนกข้อมูลด้วยวิธีเรandomฟอว์เรสส์ให้ค่าความถูกต้องสูงที่สุดจากเทคนิคทั้งหมด 3 เทคนิคอย่างมีนัยสำคัญทางสถิติที่ระดับค่าความเชื่อมั่น .05 ดังตารางที่ 3

ตารางที่ 3 ตารางการทดสอบค่าทางสถิติแบบ Paired T-Test

ที่	Algorithms	Accuracy (%)	SD.
1	Random Forest	84.29	0.44
2	Decision Tree	81.91	0.42
3	Bagging	81.71	0.45

จากตารางที่ 3 พบว่าการทดสอบจากข้อมูลชุดเดียวกัน ด้วยเทคนิคทั้ง 3 โดยให้เทคนิคการจำแนกข้อมูลด้วยวิธีเรandomฟอว์เรสส์เป็นฐานจะได้ค่าทางสถิติที่ดีกว่าเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ และเทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิง

3. ผลการทดสอบพยากรณ์ด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีเรandomฟอว์เรสส์

การพยากรณ์ด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีเรandomฟอว์เรสส์โดยการแบ่งชุดข้อมูลออกเป็น 10 ส่วน (10-fold) ผลลัพธ์ที่ได้ตามตาราง Confusion matrix ดังตารางที่ 4

ตารางที่ 4 ตารางผลการพยากรณ์เทคนิคการจำแนกข้อมูลด้วยวิธีเรandomฟอว์เรสส์

		Prediction			Recall (%)
		A	B	C	
Actual	A	16,404	794	3,074	80.9
	B	198	14,065	705	94.0
	C	3,733	1,758	24,604	81.8
Precision (%)		80.7	84.6	86.7	

โดยการแบ่งประเภทกลุ่มอาชีพจากข้อมูลภาวะการมีงานทำ ประกอบด้วย

A คือข้าราชการ/เจ้าหน้าที่หน่วยงานของรัฐ

B คือดำเนินธุรกิจอิสระ/เจ้าของกิจการ

C คือพนักงานบริษัท/องค์กรธุรกิจเอกชน

จากตารางที่ 4 พบว่าผลของการทำนายคลาสราขการ/เจ้าหน้าที่หน่วยงานของรัฐ เป็น 16,404 คลาสดำเนินธุรกิจอิสระ/เจ้าของกิจการเป็น 14,065 และคลาสพนักงานบริษัท/องค์กรธุรกิจเอกชนเป็น 24,604 รวมเป็น 55,073 จากจำนวนข้อมูลทั้งหมด 65,335 ค่าสัดส่วนข้อมูลที่ถูกต้อง (Precision) กลุ่มข้อมูลพนักงานบริษัท/องค์กรเอกชน มีค่าความแม่นยำสูงสุดที่ร้อยละ 86.7 และค่าสัดส่วนของข้อมูลที่ตรงตามความต้องการที่ถูกค้นคืนกับข้อมูลที่ต้องการทั้งหมด (Recall) กลุ่มงานธุรกิจอิสระ/เจ้าของกิจการ มีค่าสูงสุดที่ร้อยละ 94.0

อภิปรายผลการวิจัย

งานวิจัยนี้เป็นการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลเพื่อการเปรียบเทียบตัวแบบเทคนิคการพยากรณ์อาชีพ ของนักศึกษาที่สำเร็จศึกษาระดับปริญญาตรีซึ่งข้อมูลที่นำมาใช้ในงานวิจัยนี้เป็นข้อมูลภาวะการมีงานทำจากสำนักงานคณะกรรมการการอุดมศึกษา งานวิจัยนี้มีจุดเด่นจากข้อมูลที่ได้มาจากสาขาที่เกี่ยวข้องกับคอมพิวเตอร์ทุกมหาวิทยาลัยที่สังกัดอุดมศึกษา ผลการวิจัยพบว่าค่าความแม่นยำในการจำแนกประเภทข้อมูลเฉลี่ยจาก 3 เทคนิคตัวแบบการพยากรณ์อาชีพด้วยเทคนิคการจำแนกข้อมูลวิธีแรนดอมฟอรัลเรสให้มีประสิทธิภาพในการจำแนกสูงสุดด้วยค่าเฉลี่ยร้อยละ 84.29 ตามด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจให้ค่าความถูกต้องร้อยละ 81.91 และเทคนิคการจำแนกข้อมูลวิธีแบ็กกิงให้ค่าความถูกต้องร้อยละ 81.71 มีความคลาดเคลื่อนเฉลี่ยกำลังสองเท่ากับ 0.27, 0.29 และ 0.29 มีค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์เท่ากับ 0.14, 0.16 และ 0.17 ตามลำดับจากผลการวิจัยสามารถสรุปได้ว่าตัวแบบเทคนิคการจำแนกข้อมูลวิธีแรนดอมฟอรัลเรสเป็นตัวแบบที่เหมาะสมที่จะนำไปพัฒนาเป็นระบบการแนะแนวอาชีพให้กับนักศึกษาระดับปริญญาตรีสาขาคอมพิวเตอร์ต่อไป

ข้อเสนอแนะ

งานวิจัยนี้ได้นำเสนอตัวแบบเทคนิคการพยากรณ์เป็นเพียงการใช้เทคนิคส่วนหนึ่งของการทำเหมืองข้อมูลจากหลากหลายเทคนิคที่มีอยู่ในปัจจุบัน โดยนำข้อมูลภาวะการมีงานทำของสาขาวิชาทางคอมพิวเตอร์ ซึ่งมีจำนวนของข้อมูลที่ยังไม่มาก ในอนาคตหากมีการพัฒนางานวิจัยให้มีประสิทธิภาพมากขึ้นอาจจะต้องเพิ่มข้อมูลทางด้านสาขาวิชาอื่น ๆ เพื่อให้เกิดความหลากหลายของข้อมูลมากขึ้น และเพื่อให้ตัวแบบการพยากรณ์สามารถนำไปใช้ได้อย่างมีค่าความถูกต้องสูงที่สุด

เอกสารอ้างอิง

- ซัชชฎา วันดี. 2557. การศึกษาปัจจัยที่มีผลต่อการเลือกอาชีพของนิสิตระดับปริญญาตรี หลังสำเร็จการศึกษาโดยใช้เทคนิคเหมืองข้อมูล. [ปริญญาานิพนธ์]. [มหาสารคาม] : มหาวิทยาลัยมหาสารคาม.
- รัชดา ประดับศรี. 2552. การสร้างแบบประเมินสมรรถนะในการแนะแนวของเจ้าหน้าที่แนะแนวอาชีพ. [ปริญญาานิพนธ์]. [กรุงเทพฯ] : มหาวิทยาลัยศรีนครินทรวิโรฒ.
- สมฤทัย กลัดแก้ว. มปป. ระบบสนับสนุนการตัดสินใจการเลือกตำแหน่งงานให้สอดคล้องกับความสามารถของบัณฑิต. [ปริญญาานิพนธ์]. [กรุงเทพฯ] : มหาวิทยาลัยรามคำแหง.
- สุรพงศ์ เอื้อวัฒนามงคล. 2559. การทำเหมืองข้อมูล. กรุงเทพมหานคร : สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- สายชล สีนสมบูรณ์ทอง. 2558. การทำเหมืองข้อมูล. กรุงเทพมหานคร : จามจุรีโปรดักท์.
- เอกสิทธิ์ พิชรวงศ์ศักดิ์ดา. 2560. Advanced Predictive Modeling with R & Rapidminer Studio 7. กรุงเทพมหานคร : เอเชีย ดิจิตอลการพิมพ์.
- เอกสิทธิ์ พิชรวงศ์ศักดิ์ดา. 2557. การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมน์นิ่งเบื้องต้น. กรุงเทพมหานคร : ดาต้าคิวบ์.
- Ansari, G. A. 2017. Career Guidance through Multilevel Expert System Using Data Mining Technique. International Journal of Information Technology and Computer Science 9(8) : 22-29.
- Gorad, N. Zalte, J. Nandi, A. and Nayak, D. 2017. Career Counselling Using Data Mining. International Journal of Innovative Research in Computer and Communication Engineering 5(4).
- Quinlan, R. 1993. C4.5 : Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA.
- Breiman, L. and Forests, R. 2001. Machine Learning 45(1) : 5-32.
- Sathyavathi, S. Niraimathi, N. and Priyadarshini, K. 2016. Data Mining for Building an Informed Decision Making Model for Career Prediction. International Journal of Research in Computer Science 3(3) : 8-12.
- Campagni, R. Merlini, D. Sprugnoli, R. and Verri, M.C. 2015. Data mining models for student careers. Expert Systems with Applications 42(13) : 5508-5521.