

การศึกษาเทคนิคพยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีสาขาคอมพิวเตอร์ โดยใช้เทคนิคเหมืองข้อมูล

A Study of Techniques in Predicting Career Counseling for Undergraduate Students of the Computer Program by Using Data Mining Technique

สำราญ วานนท์^{1*} ธรัช อารีราษฎร์² และ จริญญา แสนราช³

สาขาวิชาการจัดการเทคโนโลยี คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม^{1,2}
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ³

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อศึกษาเทคนิคพยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูลที่เหมาะสม 2) เพื่อเปรียบเทียบผลการวิเคราะห์พยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูล งานวิจัยนี้ได้ นำข้อมูลภาวะการมีงานทำของบัณฑิต และข้อมูลระเบียบประวัติของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษา จากสำนักงานคณะกรรมการการอุดมศึกษาย้อนหลัง 5 ปี คือปี 2555 – 2559 จำนวน 65,335 ระเบียบ ในสาขาวิชาทางด้านคอมพิวเตอร์ และมีคุณลักษณะประกอบด้วย ผลการเรียนรู้ ความสามารถพิเศษ อาชีพของบิดามารดา รายได้ของบิดามารดา เพศ ตำแหน่งงาน ความสอดคล้องสาขา สาขาวิชา ทดลองวัดความแม่นยำด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอรัล และเทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิง ผลการวิจัยพบว่า ความแม่นยำในการจำแนกประเภทข้อมูล จาก 3 เทคนิค 1) เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจเท่ากับ 81.91% 2) เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอรัลเท่ากับ 84.29% และ 3) เทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิงเท่ากับ 81.71% พบว่าเทคนิคแรนดอมฟอรัลให้ความถูกต้องในการจำแนกประเภทข้อมูลสูงที่สุด

คำสำคัญ: การแนะนำอาชีพ, เหมืองข้อมูล, ต้นไม้ตัดสินใจ, แรนดอมฟอรัล, แบ็กกิง

ABSTRACT

The major aim of the research was to study techniques in predicting career counseling of undergraduate students of the computer program by using appropriate data mining technique. The study results provided data about the employment and personnel record of 65,335 graduates in the computer program under Office of the Higher Education Commission who graduated during 2012-2016. The data of graduates consisted of academic record, special abilities, parent's career, income of parents, gender, work position relating to major field.

The study showed that 1) the appropriate data mining technique consisted of techniques for data classification by using random forest method tree method for decision making, and Bagging method. 2 The findings indicated that the precision of three techniques: random forest method, tree method and Bagging method was 84.29%, 81.91%, and 81.71% respectively. The study revealed that the most precision in data classification was the random forest method.

Keyword: Career Counseling, Data Mining, Tree-Decision Making, Random Forest, Bagging

บทนำ

ปัจจุบันความต้องการจ้างงานในตลาดแรงงานของประเทศไทย ขึ้นอยู่กับสภาพการเจริญเติบโตของเศรษฐกิจ ณ ช่วงเวลานั้นๆ ทำให้การเลือกอาชีพนับว่าเป็นเรื่องสำคัญอย่างยิ่งในชีวิตมนุษย์ ต้องมีการเริ่มต้นด้วยการวางแผนตั้งแต่วัยเรียน ในสถาบันอุดมศึกษา เพื่อให้เพิ่มโอกาสในการเลือกอาชีพที่เหมาะสมกับความต้องการสอดคล้องกับความสามารถ สถาบันอุดมศึกษาจึงต้องผลิตบัณฑิตที่มีความรู้ความสามารถครบทุกด้าน พร้อมกับการส่งเสริมแนะแนวทางการอาชีพให้กับนักศึกษา แต่ปัจจุบันการแนะแนวอาชีพในสถาบันอุดมศึกษาในประเทศไทย ที่ได้ดำเนินการกันอยู่ในปัจจุบันก็เป็นการให้ความช่วยเหลือเกี่ยวกับการให้คำปรึกษาทางด้านอาชีพ อาทิเช่น การเลือกอาชีพ เลือกสถานทำงาน วิธีสมัครงาน โดยนักศึกษาจะมาขอคำปรึกษาและความช่วยเหลือจากอาจารย์ที่ทำหน้าที่ให้คำปรึกษาประจำศูนย์ให้คำปรึกษา และแนะแนวของมหาวิทยาลัยและวิทยาลัยนั้น สำหรับบางมหาวิทยาลัยได้แยกบริการจัดการงานให้นักศึกษาทำทั้งงานพิเศษและงานถาวรมาไว้ที่ฝ่ายกิจการนิสิตนักศึกษาของมหาวิทยาลัยและวิทยาลัย แต่อย่างไรก็ตาม การแนะแนวอาชีพในสถาบันอุดมศึกษาของประเทศไทยก็ยังไม่ได้ผลเต็มที่ [1] นักศึกษาจำนวนมากประสบปัญหาในการตัดสินใจเลือกอาชีพให้ตรงกับตัวเอง มีเพียงจำนวนน้อยคนที่ตัดสินใจเลือกอาชีพที่ตรงกับตัวเอง การเลือกอาชีพผิดอาจทำให้การดำรงชีวิตยุ่งยาก [2] ซึ่งอาจจะส่งผลต่อความไม่พอใจในการทำงาน ปฏิบัติหน้าที่ได้ไม่ดี เกิดความวิตกกังวล เกิดความเครียดและไม่ใส่ใจต่อสังคม [3]

ผู้วิจัยจึงมีความสนใจที่จะหาวิธีแก้ปัญหาดังกล่าว โดยการนำข้อมูลภาวะการมีงานทำกับข้อมูลระเบียบประวัติของผู้สำเร็จการศึกษาตั้งแต่ปี พ.ศ. 2555-2559 จำนวน 65,335 ระเบียบ ในสาขาวิชาทางด้านคอมพิวเตอร์ของมหาวิทยาลัยในประเทศไทย มาศึกษาเทคนิคการพยากรณ์อาชีพโดยใช้เทคนิคเหมืองข้อมูล (Data Mining) ในการจำแนกข้อมูล (Classification) เพื่อเป็นการเพิ่มความแม่นยำในการจำแนกข้อมูลให้มีความถูกต้องมากที่สุด จึงจำเป็นต้องเลือกใช้เทคนิควิธีจำแนกข้อมูลที่เหมาะสมกับข้อมูลและได้ค่าความแม่นยำที่อยู่ในระดับที่ยอมรับได้ด้วยการใช้คุณลักษณะทางด้านความสามารถหรือผลการเรียนของผู้สำเร็จการศึกษาที่ได้รับจากหลักสูตร ด้านความสามารถพิเศษ ด้านอาชีพของบิดามารดา ด้านรายได้ของบิดามารดา ด้านเพศ ด้านตำแหน่งงาน ด้านความสอดคล้องสาขา ด้านสาขาวิชา งานวิจัยนี้ได้เสนอเทคนิคการจำแนกข้อมูล 3 เทคนิคได้แก่ เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอเรส (Random Forest) และเทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิง (Bagging) สำหรับวิเคราะห์ค่าความแม่นยำในการพยากรณ์ และนำเทคนิคที่ให้ค่าความแม่นยำมากที่สุดไปใช้เป็นตัวแบบในการพัฒนาระบบพยากรณ์เพื่อการแนะแนวอาชีพต่อไป

1. วัตถุประสงค์การวิจัย

- 1) เพื่อศึกษาเทคนิคพยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูลที่เหมาะสม
- 2) เพื่อเปรียบเทียบผลการวิเคราะห์พยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูล

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

เหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูลเป็นกระบวนการค้นหาความสัมพันธ์ รูปแบบและแนวโน้มใหม่ ๆ โดยการใช้วิธีการทางคณิตศาสตร์และสถิติวิเคราะห์ข้อมูล [4] เพื่อให้ได้ผลลัพธ์ที่มีความถูกต้องสมบูรณ์คือ ความรู้ที่แฝงอยู่ในข้อมูลที่สามารถนำไปใช้ในการตัดสินใจและดำเนินงานโดยไม่เกิด ความผิดพลาดหรือสร้างความเสียหายในภายหลัง [9] โดยมีขั้นตอนต่าง ๆ ตามมาตรฐาน 6 ขั้นตอน (CRISP-DM) 1) ขั้นตอนการทำความเข้าใจการวิจัยและธุรกิจ

(Business/Research Understanding) 2) ขั้นตอนการทำความเข้าใจข้อมูล (Data Understanding) 3) ขั้นตอนการจัดเตรียมข้อมูล (Data Preparation) 4) ขั้นตอนการสร้างตัวแบบ (Modeling) 5) ขั้นตอนการประเมินตัวแบบ (Evaluation) 6) ขั้นตอนการใช้งาน (Deployment)

เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree)

การสร้างโมเดลด้วยต้นไม้ตัดสินใจ [5] จะทำการคัดเลือกคุณลักษณะที่มีความสัมพันธ์กับคลาสมากที่สุดขึ้นมาเป็นโหนดบนสุดของ Tree (root node) หลังจากนั้นก็จะหาคุณลักษณะถัดไปเรื่อย ๆ ในการหาความสัมพันธ์ของคุณลักษณะนี้จะใช้ตัววัดที่เรียกว่า Information Gain (IG) การคำนวณค่า Information Gain จะใช้ค่า Entropy ซึ่งเป็นการวัดความแตกต่างหรือการกระจายของข้อมูลถ้าข้อมูลมีความแตกต่างกันมากค่า Entropy ก็จะมีค่าสูง ในทางตรงข้ามถ้าข้อมูลมีความคล้ายกันมากค่า Entropy ก็จะมีค่าต่ำ อัลกอริทึมของเทคนิคต้นไม้ตัดสินใจส่วนใหญ่ไม่รองรับข้อมูลแบบต่อเนื่อง จึงต้องมีการแบ่งให้เป็นข้อมูลแบบไม่ต่อเนื่องเสียก่อน อัลกอริทึมสำหรับการแบ่งให้ข้อมูลเป็นแบบไม่ต่อเนื่องคือ ID3, C4.5 และ C5.0 เป็นต้น สำหรับงานวิจัยนี้ใช้อัลกอริทึมของเทคนิคต้นไม้ตัดสินใจ C4.5 [6] โดยการพิจารณาค่าบ่งชี้ความเหมาะสมของ Attribute เรียกว่า Gini (Gini Index) และคำนวณหาค่าตัวบ่งชี้ในการจำแนกประเภทของชุดข้อมูล หรือ Gini-split โดย attribute ที่มีค่าน้อยที่สุดถือว่ามีความสามารถในการจำแนกประเภทได้ดีที่สุด

เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอเรส (Random Forest)

การจำแนกข้อมูลด้วยวิธีแรนดอมฟอเรสเป็นชุดของการจำแนกประเภทแบบไม่ตัดแต่งกิ่ง (unpruned) หรือต้นไม้ถดถอย (Regression Trees) ซึ่งถูกสร้างจากการนำข้อมูลฝึกสอนไป สุ่มเลือกตัวอย่างข้อมูลและคุณลักษณะข้อมูลแล้วนำมาสร้างเป็นต้นไม้ตัดสินใจซึ่งมีตัวอย่างส่วนหนึ่งที่ไม่ถูกเลือกเรียกข้อมูลส่วนนี้ว่า Out-of-Bag (OOB) จะถูกนำมาใช้ในการทดสอบต้นไม้ตัดสินใจ [6] และนำผลการทำนายที่ได้มาโหวต (Vote) เพื่อหาคำตอบสุดท้าย [8]

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 [9]

2.2 งานวิจัยที่เกี่ยวข้อง

ชัชชญา วันดี [10] ศึกษาวิจัยเรื่อง เปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลการเลือกอาชีพของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษา โดยในงานวิจัยนี้ได้ใช้ชุดข้อมูลภาวะการมีงานทำของบัณฑิต และข้อมูลประวัติของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษา คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม ระหว่างปี พ.ศ. 2550-2554 จำนวน 12 คุณลักษณะ และ 2,515 ระเบียบ ซึ่งได้นำเทคนิคแบบจำลองต้นไม้ตัดสินใจ (Decision Tree) เทคนิคโครงข่ายประสาทเทียม (Neural Network) และเทคนิคการเรียนรู้แบบเบย์ (Naive Bayes) มาทำการเปรียบเทียบประสิทธิภาพ ผลจากการศึกษาพบว่าประสิทธิภาพในการจำแนกข้อมูลแบบต้นไม้ตัดสินใจ มีประสิทธิภาพในการจำแนกสูงสุดด้วยค่าเฉลี่ย 88.62% และปัจจัยสำคัญที่ทำให้การเลือกอาชีพตรงหรือไม่ตรงกับสาขา มี 4 ปัจจัย คือสาขาวิชาที่เรียน เกรดเฉลี่ยเฉพาะวิชาสาขา เพศ และเกรดเฉลี่ยรวม

สมฤทัย กลัดแก้ว [11] การศึกษามีวัตถุประสงค์ 1) เพื่อศึกษาปัจจัยที่มีผลต่อการเลือกตำแหน่งงานให้สอดคล้องกับความสามารถของบัณฑิตด้วยการวิเคราะห์การถดถอยโลจิสติกพหุกลุ่มและการจำแนกประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจ 2) เพื่อเปรียบเทียบความถูกต้องของการพยากรณ์ระหว่างการวิเคราะห์การถดถอยโลจิสติกพหุกลุ่ม (Logistic Regression) และการจำแนกประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจ 3) เพื่อพัฒนาตัวแบบการตัดสินใจการเลือกตำแหน่งงานให้สอดคล้องกับความสามารถของบัณฑิต กลุ่มตัวอย่างที่ใช้ในการศึกษาได้แก่ ข้อมูลภาวะบัณฑิตที่มีงานทำที่เข้ารับพระราชทานปริญญาบัตรในปีการศึกษา 2555-2557 จำนวน 1,933 คน เครื่องมือที่ใช้ในการศึกษา ได้แก่ แบบสำรวจข้อมูลภาวะบัณฑิตมีงานทำของศูนย์คอมพิวเตอร์มหาวิทยาลัยราชภัฏหมู่บ้านจอมบึง สถิติที่ใช้ในการวิเคราะห์ข้อมูล ได้แก่ การวิเคราะห์การถดถอยโลจิสติกพหุกลุ่มและการจำแนกประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจพบว่าเทคนิคการจำแนกประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจมีความถูกต้องมากกว่าการวิเคราะห์การถดถอยโลจิสติกพหุกลุ่มเล็กน้อย โดยค่าความถูกต้องของเทคนิคการจำแนกประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจ เท่ากับ 57.37% และการวิเคราะห์การถดถอยโลจิสติกพหุกลุ่มมีความถูกต้อง 56.3%

Sathyavathi, Niraimathi, and Priyadarshini [12] ศึกษาวิจัยเรื่อง เหมืองข้อมูลสำหรับสร้างตัวแบบข้อมูลเพื่อการตัดสินใจสำหรับการพยากรณ์อาชีพ ผลการวิจัยพบว่าปัจจัยที่มีผลต่อการตัดสินใจในการเลือกอาชีพโดยการใช้เครื่องมือเหมืองข้อมูลแบบจำแนกกลุ่ม ID3, CHAID โดยใช้ขั้นตอนในการดำเนินการทำเหมืองข้อมูลแบบจำแนกกลุ่ม ใช้โปรแกรม Rapid miner เป็นเครื่องมือ ขั้นตอนวิธีที่ใช้สำหรับการเปรียบเทียบในการจัดกลุ่มข้อมูลคือ ID3, CHAID, Decision tree ประสิทธิภาพที่ได้ค่าสูงสุดเมื่อเปรียบเทียบจากขั้นตอนวิธีดังกล่าวจะพบว่า Decision tree ได้ให้ค่าความถูกต้องสูงที่สุดในการพยากรณ์สำหรับการจำแนกกลุ่มข้อมูล

Campagni, Merlini, Sprugnoli and Verri [13] ศึกษาวิจัยเรื่อง ตัวแบบการทำเหมืองสำหรับอาชีพนักศึกษา ผลการวิจัยพบว่าระเบียบวิธีการทำเหมืองข้อมูลเพื่อวิเคราะห์อาชีพของนักศึกษาในมหาวิทยาลัย ความแตกต่างทางวิธีการจัดกลุ่มและลำดับรูปแบบของเทคนิคการจำแนก กลวิธีในการปรับปรุงประสิทธิภาพของนักศึกษาตามตารางการสอบ โดยจะแนะนำถึงอาชีพในอนาคตก่อน นักศึกษาจะต้องทำการทดสอบหลังจากเรียนจบให้สอดคล้องกับหลักสูตรที่เรียนโดยทันที แล้วเปรียบเทียบกับนักศึกษาทั่วไปกับนักศึกษาที่ใช้เทคนิคที่มีความแตกต่างกันและประยุกต์ใช้วิธีการดังกล่าวกับนักศึกษาจริง พบว่านักศึกษาที่ได้รับแนวคิดอาชีพในอนาคตจะมีผลการเรียนในเทอมและจบการศึกษาที่ดีขึ้น

วิธีดำเนินการวิจัย

1. ขั้นตอนการดำเนินการวิจัยประกอบด้วย

1. ศึกษาเทคนิคการพยากรณ์อาชีพที่เป็นไปได้และเหมาะสมในการนำมาพยากรณ์อาชีพจากเอกสารและงานวิจัยที่เกี่ยวข้องและสรุปผลเลือกเทคนิคที่เหมาะสม

2. จัดเตรียมข้อมูล โดยประกอบด้วยข้อมูลภาวะการมีงานทำกับข้อมูลระเบียบประวัติของผู้สำเร็จการศึกษาตั้งแต่ปี พ.ศ. 2555 - 2559 จำนวน 65,335 ระเบียบ ในสาขาวิชาทางด้านคอมพิวเตอร์ ของมหาวิทยาลัยในประเทศไทย เพื่อนำมาวิเคราะห์ด้วยเทคนิคที่เลือกไว้ในขั้นตอนที่ 1. โดยมีขั้นตอนตามมาตรฐาน 6 ขั้นตอนดังนี้

1) ทำความเข้าใจเกี่ยวกับภาวะการมีงานทำของบัณฑิต โดยศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องเทคนิคการพยากรณ์ความสัมพันธ์ข้อมูลการแนะนำอาชีพสำหรับนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูล

2) ทำความเข้าใจข้อมูล ข้อมูลที่นำมาใช้ในการวิจัยในครั้งนี้ เป็นข้อมูลภาวะการมีงานทำของบัณฑิตและระเบียบประวัติผลการเรียนของนักศึกษา ประกอบไปด้วยคุณลักษณะ เพศ อาชีพบิดา รายได้บิดา อาชีพมารดา รายได้มารดา ผลการเรียนเฉลี่ย สาขาวิชา ความสามารถพิเศษ ความสอดคล้องของอาชีพกับสาขาวิชา และประเภทตำแหน่งงาน แสดงดังตารางที่ 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) นำข้อมูลจากแต่ละตารางมาทำการรวมกันโดยข้อมูลจะประกอบด้วยข้อมูลระบุประวัติและข้อมูลภาวะการมีงานทำ

3.2) จัดรูปแบบข้อมูล (Data Transformation) ทำการจัดรูปแบบข้อมูลเพื่อให้ข้อมูลมีความพร้อมที่จะนำไปใช้งานต่อไป

3.3) การแบ่งส่วนข้อมูล การแบ่งข้อมูลออกเป็นส่วนหลายส่วนเท่าๆ กันแล้วนำข้อมูลที่แบ่งไว้ไปใช้ในการสร้างตัวแบบและการทดสอบตัวแบบ เวียนกันไปจนครบทุกส่วนของข้อมูล การกำหนดจำนวนส่วนของข้อมูลแบ่งออกเป็น 10 ส่วน (10 folds Cross-Validation)

4) การวิเคราะห์ข้อมูลและสร้างตัวแบบโดยการใช้เทคนิคที่ศึกษาโดยมีเทคนิคที่เลือก ด้วยโปรแกรม Weka

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 ได้แก่

สถิติการพยากรณ์ โดยพิจารณาค่าความถูกต้องที่มีค่าสูงสุด (Accuracy) ความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) และค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ (MAE) ที่มีค่าต่ำที่สุด

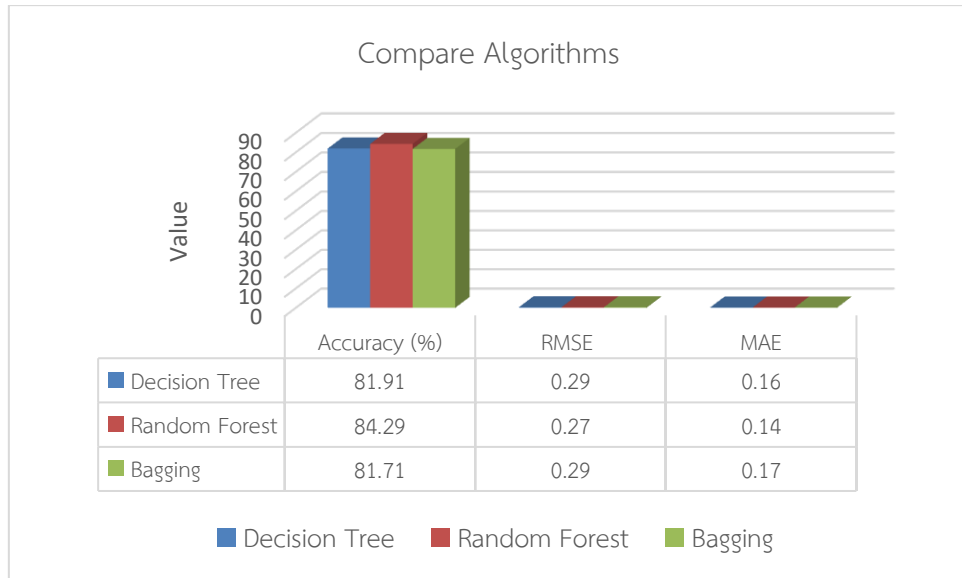
ผลการวิจัย

1. ผลการศึกษาเทคนิคการพยากรณ์จำแนกข้อมูลโดยใช้เทคนิคเหมืองข้อมูล

ผู้วิจัยได้ศึกษาเทคนิคเหมืองข้อมูลการจำแนกข้อมูล โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอรัล เทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิง และเทคนิคการจำแนกข้อมูลด้วยวิธีบูสต์ติง วิเคราะห์เปรียบเทียบกับข้อมูลภาวะการมีงานทำใช้ค่าความถูกต้อง ความคลาดเคลื่อนเฉลี่ยกำลังสอง ค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ เพื่อวัดประสิทธิภาพของตัวแบบ ได้ผลการดำเนินงาน ดังตารางที่ 2 และภาพที่ 1

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 1 แสดงกราฟการเปรียบเทียบผลการวิเคราะห์ค่าความถูกต้องของแต่ละเทคนิค

จากตารางที่ 2 พบว่าตัวแบบการพยากรณ์อาชีพด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอเรสให้ค่าความถูกต้องสูงที่สุดร้อยละ 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

จากการเปรียบเทียบผลการวิเคราะห์ค่าความถูกต้องของแต่ละเทคนิคสร้างตัวแบบการพยากรณ์อาชีพจากข้อมูลภาวะการมีงานทำของบัณฑิต นำมาทดสอบค่าทางสถิติแบบ Paired T-Test พบว่าเทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอเรสให้ค่าความถูกต้องสูงที่สุดจากเทคนิคทั้งหมด 3 เทคนิคอย่างมีนัยสำคัญทางสถิติที่ระดับค่าความเชื่อมั่น .05 ดังตารางที่ 3

ตารางที่ 3 ตารางการทดสอบค่าทางสถิติแบบ Paired T-Test

ที่	Algorithms	Accuracy (%)	SD.	Number of Time
1	Random Forest	84.29	0.44	V/ /*
2	Decision Tree	81.91	0.42	0/0/1
3	Bagging	81.71	0.45	0/0/1

จากตารางที่ 3 พบว่าการทดสอบจากข้อมูลชุดเดียวกัน ด้วยเทคนิคทั้ง 3 โดยให้เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอร์เรสเป็นฐานจะได้ค่าทางสถิติที่ดีกว่าเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ และเทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิง

3. ผลการทดสอบพยากรณ์ด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอร์เรส

การพยากรณ์ด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอร์เรส โดยการแบ่งชุดข้อมูลออกเป็น 10 ส่วน (10-fold) ผลลัพธ์ที่ได้ตามตาราง confusion matrix ที่ 3

ตารางที่ 3 ตารางผลการพยากรณ์เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอร์เรส

		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	

โดยที่

A คือ ข้าราชการ /เจ้าหน้าที่หน่วยงานของรัฐ

B คือ ดำเนินธุรกิจอิสระ/เจ้าของกิจการ

C คือ พนักงานบริษัท/องค์กรธุรกิจเอกชน

จากตาราง Confusion Matrix ที่ 3 ผลของการทำนายคลาสราชการ/เจ้าหน้าที่หน่วยงานของรัฐเป็น 16,404 คลาสดำเนินธุรกิจอิสระ/เจ้าของกิจการเป็น 14,065 และคลาพนักงานบริษัท/องค์กรธุรกิจเอกชนเป็น 24,604 รวมเป็น 55,073 จากจำนวนข้อมูลทั้งหมด 65,335 ให้ค่าความถูกต้องเป็น 84.29 %

อภิปรายผลการวิจัย

งานวิจัยนี้เป็นการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลเพื่อการเปรียบเทียบตัวแบบเทคนิคการพยากรณ์อาชีพ ของนักศึกษาที่สำเร็จศึกษาระดับปริญญาตรีซึ่งข้อมูลที่น่ามาใช้ในการงานวิจัยนี้เป็นข้อมูลภาวะการมีงานทำจากสำนักงานคณะกรรมการการอุดมศึกษา จากผลการวิจัยพบว่าความแม่นยำในการจำแนกประเภทข้อมูลเฉลี่ยจาก 3 เทคนิค ตัวแบบการพยากรณ์อาชีพด้วยเทคนิคการจำแนกข้อมูลวิธีแรนดอมฟอร์เรสให้ค่าความถูกต้องสูงที่สุดร้อยละ 84.29 ตามด้วยเทคนิคการจำแนกข้อมูลวิธีต้นไม้ตัดสินใจให้ค่าความถูกต้องร้อยละ 81.91 และเทคนิคการจำแนกข้อมูลวิธีแบ็กกิงให้ค่าความถูกต้องร้อยละ 81.71 มีความคลาดเคลื่อนเฉลี่ยกำลังสองเท่ากับ 0.27, 0.29 และ 0.29 มีค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์เท่ากับ 0.14, 0.16 และ 0.17 ตามลำดับ จากผลการวิจัยสามารถสรุปได้ว่าตัวแบบเทคนิคการจำแนกข้อมูลวิธีแรนดอมฟอร์เรสเป็นตัวแบบที่เหมาะสมที่จะนำไปพัฒนาเป็นระบบการแนะแนวอาชีพให้กับนักศึกษาระดับปริญญาตรีสาขาคอมพิวเตอร์ต่อไป

ข้อเสนอแนะ

งานวิจัยนี้ได้นำเสนอตัวแบบเทคนิคการพยากรณ์เป็นเพียงการใช้เทคนิคส่วนหนึ่งของการทำเหมืองข้อมูล จากหลากหลายเทคนิคที่มีอยู่ในปัจจุบัน โดยนำข้อมูลภาวะการมีงานทำของสาขาวิชาทางคอมพิวเตอร์ ซึ่งมีจำนวน ของข้อมูลที่ยังไม่มาก ในอนาคตหากมีการพัฒนางานวิจัยให้มีประสิทธิภาพมากขึ้นอาจจะต้องเพิ่มข้อมูลทางด้าน สาขาวิชาอื่นๆ เพื่อให้เกิดความหลากหลายของข้อมูลมากขึ้น และเพื่อให้ตัวแบบการพยากรณ์สามารถนำไปใช้ได้ อย่างมีค่าความถูกต้องสูงที่สุด

เอกสารอ้างอิง

- [1] รัชดา ประดับศรี. (2552). การสร้างแบบประเมินสมรรถนะในการแนะแนวของเจ้าหน้าที่แนะแนวอาชีพ. ปรินญาณิพนธ์ การศึกษามหาบัณฑิต สาขาจิตวิทยาการแนะแนว มหาวิทยาลัยศรีนครินทรวิโรฒ.
- [2] Ansari, Gufran Ahmad. (2017). Career Guidance through Multilevel Expert System Using Data Mining Technique. *International Journal of Information Technology and Computer Science*, 9(8), 22-29. Retrieved from <http://www.mecs-press.org/ijitcs/v9-n8/v9n8-3.html>
- [3] Nikita Gorad. Ishani Zalte. Aishwarya Nandi. & Deepali Nayak. (2017, April). Career Counselling Using Data Mining. *International Journal of Innovative Research in Computer and Communication Engineering*, 5(4). Retrieved from <http://www.rroj.com/open-access/career-counselling-using-data-mining-.pdf>
- [4] สายชล สีนสมบูรณ์ทอง. (2558). การทำเหมืองข้อมูล. กรุงเทพมหานคร: จามจุรีโปรดักท์.
- [5] เอกสิทธิ์ พัชรวงศ์ศักดิ์. (2557). การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไม่นิ่ง เบื้องต้น. กรุงเทพมหานคร: ดาต้าคิวบ์.
- [6] Ross Quinlan. (1993). C4.5: *Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA.
- [7] L.Breiman and R. Forests. (2001). *Machine Learning*. 45(1),5-32. Retrieved from <http://dx.doi.org/10.1023%2FA%3A1010933404324>
- [8] เอกสิทธิ์ พัชรวงศ์ศักดิ์. (2560). *Advanced Predictive Modeling with R & Rapidminer Studio 7*. กรุงเทพมหานคร: เอเชีย ดิจิตอลการพิมพ์.
- [9] สุรพงศ์ เอื้อวัฒนามงคล. (2559). การทำเหมืองข้อมูล. กรุงเทพมหานคร: สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- [10] ชัชชญา วันดี. (2557). การศึกษาปัจจัยที่มีผลต่อการเลือกอาชีพของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษา โดยใช้เทคนิค เหมืองข้อมูล. (ปรินญาณิพนธ์บัณฑิต), มหาวิทยาลัยมหาสารคาม, มหาสารคาม.
- [11] สมฤทัย กลัดแก้ว. (มปป). ระบบสนับสนุนการตัดสินใจการเลือกตำแหน่งงานให้สอดคล้องกับความสามารถของบัณฑิต. (ปรินญาณิพนธ์บัณฑิต), มหาวิทยาลัยรามคำแหง, กรุงเทพฯ.
- [12] 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. Retrieved from <http://researchscrip.com/wp-content/uploads/2016/04/IJRC5030303.pdf>
- [13] Campagni.R, Merlini.D , Sprugnoli.R, and Verri.M. (2015). Data mining models for student careers. *Expert Systems with Applications*, 42(13), 5508-5521. Retrieved from <http://dx.doi.org/10.1016/j.eswa.2015.02.052>