



การวิเคราะห์รูปแบบความสนใจเลือกอาชีพของผู้เรียนการแบบทดสอบ

ตามทฤษฎีของจอห์น ฮอลแลนด์ ด้วยเทคนิคเหมืองข้อมูล

Analysis on Behavioral Patterns of to enhance career test

John Holland theory by using Data Mining Technique

ปิยนันท์ คงไฟ^{1*} จิรพันธุ์ ศรีสมพันธุ์¹ และ จริญญา แสนราช¹

¹ภาควิชาคอมพิวเตอร์ศึกษา คณะครุศาสตร์อุตสาหกรรม

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ตำบลวงศ์สว่าง อำเภอบางซื่อ กรุงเทพมหานคร 10800

*E-mail: piyanan.bo@gmail.com

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์ความสนใจเลือกอาชีพของผู้เรียนจากแบบทดสอบตามทฤษฎีของจอห์น ฮอลแลนด์ ด้วยเทคนิคเหมืองข้อมูล ผู้วิจัยได้นำข้อมูลจากนักศึกษา วิทยาลัยเทคโนโลยีบริหารธุรกิจงานจํานวน 3,000 คน ซึ่งมีคุณลักษณะประกอบด้วย เพศ อายุ สาขาวิชา เกรดเฉลี่ยสะสม เป้าหมายทางการศึกษา ภูมิลาเนา อาชีพบิดามารดา และรายได้เฉลี่ยของครอบครัว แล้วทำการพัฒนาตัวแบบจำลองด้วยเทคนิคการจําแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจทั้งหมด 3 อัลกอริทึม ด้วยอัลกอริทึม J48, LMT, Random Forest เทคนิคการจําแนกข้อมูลด้วยการเรียนรู้แบบเบย์ และเทคนิคการจําแนกข้อมูลด้วยวิธีโครงข่ายประสาทเทียม โดยใช้ 10-Fold Cross Validation ในการแบ่งข้อมูลออกเป็นชุด ซึ่งประกอบด้วยชุดข้อมูลการสอนและชุดข้อมูลทดสอบ หลังจากนั้นทำการวัดค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-measure) เพื่อทำการวัดประสิทธิภาพในการพยากรณ์ของตัวแบบจำลอง ซึ่งผลการทดลองประสิทธิภาพในการพยากรณ์ตัวแบบจำลอง พบว่า เทคนิคการจําแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ด้วยอัลกอริทึม Random Forest สามารถพยากรณ์ได้ดีกว่า อัลกอริทึม J48, LMT เทคนิคการจําแนกข้อมูลด้วยการเรียนรู้แบบเบย์ และเทคนิคการจําแนกข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ซึ่งมีค่าความถูกต้องร้อยละ 91.37 ค่าความแม่นยำร้อยละ 91.30 ค่าความระลึกร้อยละ 91.40 และค่าความถ่วงดุลร้อยละ 91.30

Received: March 05, 2018

Revised: April 10, 2019

Accepted: May 02, 2019

ตามลำดับ ซึ่งผลการทดลองนี้สามารถนำไปประยุกต์ใช้กับนักศึกษาและหน่วยงานที่เกี่ยวข้องเพื่อวางแผนพัฒนาโครงสร้างหลักสูตรหรือวางแผนการศึกษาให้กับผู้เรียนในอนาคตได้

คำสำคัญ: ทฤษฎีของจอห์น ฮอลแลนด์ เทคนิคเหมืองข้อมูล วิธีต้นไม้ตัดสินใจ วิธีการเรียนรู้แบบเบย์ วิธีโครงข่ายประสาทเทียม

Abstract

The objective of this research was to analysis to enhance career test inventory based-on John Holland theory by using Data Mining Technique. The surveyed questionnaire was used with 3,000 students at Chamnong Business Administration Technological College. The data of graduates consisted of Sex Age Branch GPA Educational goals Domicile Parenting career and The average income of the family. This research used and compared the results of Decision Tree J48, Logistic Model Trees (LMT), Random Forest (RF), Naïve Bayes and Artificial Neural Network were used to build the prediction models. Moreover, 10-fold cross validation utilized to split the data into the training and test set. This research has measured performance models with accuracy precision Recall and F-measure. Experimental result demonstrated that Random Forest superior to C4.5 LMT Naïve Bayes and Artificial Neural Network with accuracy 91.37%, precision 91.30% Recall 91.40% and F-measure 91.30%. The results in this experiment can be applied to the relevant authorities to development plans course structure or education plan.

Keywords: John Holland theory, Decision tree, Naïve Bayes, Artificial Neural Network

1. บทนำ

การพัฒนาอย่างรวดเร็วของสังคมและเศรษฐกิจ ส่งผลกระทบต่อวิถีชีวิตของประชาชนในสังคมตลอดจนการจ้างงานในตลาดแรงงานของประเทศไทยขึ้นอยู่กับเศรษฐกิจเติบโตของเศรษฐกิจ ณ ช่วงเวลานั้น [1] ทำให้การส่งเสริมความสนใจในการเลือกอาชีพของผู้เรียน นับว่าเป็นเรื่องสำคัญอย่างยิ่งต้องมีการวางแผนตั้งแต่วัยเรียน [2] เพื่อให้เพิ่มโอกาสในการเลือกอาชีพที่เหมาะสมกับความต้องการสอดคล้องกับความสามารถ ความสนใจ ความถนัด ทักษะ ตลอดจนบุคลิกภาพของตน ดังจะเห็นได้ว่าคนเรามีความถนัด ความสามารถ

และความสนใจในงานอาชีพแตกต่างกัน บางคนเหมาะที่จะทำงานด้านหัตถกรรม บางคนเหมาะสมที่จะทำงานเกี่ยวกับเครื่องจักรกล หรืองานที่เกี่ยวข้องกับด้านวิทยาศาสตร์ ส่วนบางคนอาจจะเหมาะสมที่จะทำงานเกี่ยวกับการสอนดนตรี การช่วยเหลือติดต่อประชาชน เหล่านี้เป็นต้น แต่ละคนมีพรสวรรค์ไม่เหมือนกัน ไม่สามารถทำงานชนิดเดียวกันได้หมดทุกคน ข้อสำคัญก็คือ ต้องรู้จักตัวเอง และรู้จักงานอาชีพต่าง ๆ อย่างกว้างขวาง รวมทั้งพิจารณาว่ามิจานอะไรบ้างที่ชอบและสนใจมากที่สุด และงานนั้น ๆ เหมาะสมกับอุปนิสัยและบุคลิกลักษณะหรือไม่ [3] จากคำกล่าวข้างต้น

สอดคล้องกับสถานการณ์ปัจจุบัน พบว่านักศึกษาจำนวนมากประสบปัญหาในการตัดสินใจเลือกอาชีพให้ตรงกับตนเอง ซึ่งน้อยคนที่ตัดสินใจเลือกอาชีพที่ตรงกับความสามารถ ความถนัดของตัวเอง การเลือกอาชีพผิดทำให้การดำรงชีวิตยุ่งยาก [4] ซึ่งอาจจะส่งผลต่อความไม่พอใจในการทำงาน [5] เกิดความเบื่อหน่าย ไม่มีความสุขและมีโอกาสประสบความสำเร็จในการประกอบอาชีพได้ยาก [1]

จากคำกล่าวข้างต้นผู้วิจัยจึงแนวคิดว่าแบบทดสอบความสนใจทางอาชีพ โดยอ้างอิงจากทฤษฎีการจำแนกอาชีพตามบุคลิกภาพของ John L. Holland ได้ตั้งสมมติฐานที่ว่า “บุคลิกภาพของคนสะท้อนผ่านการเลือกอาชีพของตน” [6] คนที่เลือกอาชีพได้สอดคล้องกับบุคลิกภาพของตนมากที่สุด จะมีความพึงพอใจในอาชีพและส่งผลให้ประสบความสำเร็จในอาชีพนั้น ๆ ได้ [7] และนำเทคโนโลยีการทำเหมืองข้อมูล (Data Mining) มาช่วยในการจำแนกข้อมูลความสนใจเลือกอาชีพของผู้เรียน ซึ่งการทำเหมืองข้อมูลคือการกระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบ แนวทางและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์ ซึ่งในปัจจุบันได้มีการนำความรู้ [2] เกี่ยวกับการทำเหมืองข้อมูล มาใช้ในการจำแนกและการพยากรณ์ในงานด้านต่าง ๆ ไม่ว่าจะเป็นด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ทางด้านการศึกษาเพื่อจำแนกข้อมูลนักศึกษาในด้านต่าง ๆ เช่น การพยากรณ์ความสำเร็จการศึกษาของนักศึกษา การพยากรณ์อาชีพสำหรับนักศึกษา และการจำแนกปัจจัยที่มีผลต่อการเลือกอาชีพของนิสิตหลังสำเร็จการศึกษา เป็นต้น สอดคล้องกับงานวิจัย ศึกษาการเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์ความสำเร็จการศึกษาของนักเรียนระดับ

ประกาศนียบัตรวิชาชีพ ผลการทดลองพบว่าแบบจำลอง C4.5 ให้ความถูกต้อง 95.36% ซึ่งสูงกว่า Random Forest, Random Tree, REP Tree, k-NN และ SVM สอดคล้องกับงานวิจัย [9] ศึกษาเทคนิคการพยากรณ์อาชีพและการเปรียบเทียบผลการวิเคราะห์พยากรณ์อาชีพของนักศึกษา ผลการทดลองพบว่า เทคนิคการจำแนกข้อมูลด้วยวิธีเรแนคคอมฟอร์เรตให้ความถูกต้อง 84.29% ซึ่งสูงกว่าการเทคนิคต้นไม้ตัดสินใจและแบ็กกิง สอดคล้องกับงานวิจัย [8] ศึกษาปัจจัยที่มีผลต่อการเลือกอาชีพของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษา ผลการทดลองพบว่า การจำแนกข้อมูลแบบต้นไม้ตัดสินใจให้ความถูกต้อง 88.26% ซึ่งสูงกว่าเทคนิคโครงข่ายประสาทเทียมและเทคนิคการเรียนรู้แบบเบย์ จากผลวิจัยข้างต้นการเลือกใช้เทคนิควิธีที่เหมาะสมในการจำแนกข้อมูลส่งผลให้ประสิทธิภาพในการจำแนกข้อมูลมีความถูกต้องมากและเหมาะสมกับการนำไปวิเคราะห์ข้อมูล

ในการวิจัยครั้งนี้ ทำการสำรวจข้อมูลจากนักศึกษา วิทยาลัยเทคโนโลยีบริหารธุรกิจ จำนวน 3,000 คน เพื่อนำมาพัฒนาแบบจำลองที่สามารถจำแนกข้อมูลความสนใจเลือกอาชีพของผู้เรียน เพื่อเป็นแนวทางในการตัดสินใจเลือกอาชีพของผู้เรียน ซึ่งมีวัตถุประสงค์ในการวิจัยคือ เพื่อวิเคราะห์ความสนใจเลือกอาชีพของผู้เรียน โดยใช้แบบทดสอบตามทฤษฎีของจอห์น ฮอลแลนด์ ด้วยเทคนิคเหมืองข้อมูล โดยประกอบด้วยตัวแปรคุณลักษณะดังนี้ เพศ อายุ ระดับการศึกษา สาขาวิชาเกรดเฉลี่ยสะสม เป้าหมายทางการศึกษา ภูมิฐานะเดิม อาชีพบิดา อาชีพมารดา รายได้เฉลี่ยต่อเดือนของครอบครัว และกลุ่มบุคลิกภาพ โดยในงานวิจัยนี้ นำเทคนิคการจำแนกข้อมูลด้วยอัลกอริทึม 5 อัลกอริทึมได้แก่ เทคนิคการจำแนกข้อมูลด้วยวิธี

ต้นไม้ตัดสินใจ โดยใช้อัลกอริทึม J48 LMT Random Forest เทคนิคการจำแนกข้อมูลด้วยการเรียนรู้แบบเบย์ และเทคนิคการจำแนกข้อมูลด้วยวิธีโครงข่ายประสาทเทียม เพื่อทำการเปรียบเทียบหาอัลกอริทึมที่วิเคราะห์ค่าความแม่นยำในการจำแนกข้อมูล และนำเทคนิคที่ให้ค่าความแม่นยำมากที่สุดไปใช้เป็นตัวแบบจำลองในพัฒนารูปแบบความสนใจเลือกอาชีพของผู้เรียนต่อไป

2. วิธีดำเนินการวิจัย

2.1 ข้อมูลที่นำมาใช้ในการทดลอง

ข้อมูลที่นำมาใช้ในงานวิจัยนี้เป็นข้อมูลทั่วไปของนักศึกษา ข้อมูลครอบครัวและผู้ปกครอง และข้อมูลความสนใจด้านอาชีพตามทฤษฎีของจอห์น สอลแลนด์ โดยทำการจัดเก็บด้วยแบบสอบถามความสนใจการเลือกอาชีพจากนักศึกษา วิทยาลัยเทคโนโลยีบริหารธุรกิจจันทักษ์

2.2 ขั้นตอนการดำเนินการวิจัย

ประกอบด้วย 3 ขั้นตอน ดังนี้

2.2.1 ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการจำแนกข้อมูลความสนใจเลือกอาชีพของผู้เรียนโดยใช้แบบทดสอบตามทฤษฎีของจอห์น สอลแลนด์ ด้วยเทคนิคเหมือนข้อมูล เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ อัลกอริทึม J48 LMT Random Forest เทคนิคการจำแนกข้อมูลด้วยการเรียนรู้แบบเบย์ และเทคนิคการจำแนกข้อมูลด้วยวิธีโครงข่ายประสาทเทียม

2.2.2 การเตรียมข้อมูล หลังจากได้ข้อมูลทีเก็บจากผู้เรียนด้วยแบบสอบถาม ผู้วิจัยได้กำหนดคุณลักษณะประจำหรือแอททริบิวต์ต่าง ๆ ประกอบด้วย เพศ อายุ ระดับการศึกษา สาขาวิชา เกรดเฉลี่ยสะสม เป้าหมายทางการศึกษา ภูมิลำเนาเดิม อาชีพบิดา อาชีพมารดา รายได้เฉลี่ยต่อเดือน

ของครอบครัว และกลุ่มบุคลิกภาพ ทั้งหมด 11 แอททริบิวต์ ดังตารางที่ 1

2.2.2.1 การรวบรวมข้อมูล ได้ทำการจัดเก็บข้อมูลแบบสอบถามจากนักศึกษาจำนวน 3,407 ชุด แต่มีแบบสอบถามที่ตอบคำถามได้ครบถ้วนสมบูรณ์จำนวน 3,257 ชุด

2.2.2.2 การกลั่นกรองข้อมูล หลังจากขั้นตอนที่แล้วผู้วิจัยนำข้อมูลทั้งหมดมาจัดการข้อมูลที่ขาดหาย (Missing Value) และการจัดการข้อมูลรบกวน (Noisy data) ต่าง ๆ หรือที่มีค่าผิดปกติ (Outliers) รวมทั้งข้อมูลที่มีความผิดพลาดจากการพิมพ์ข้อมูล ในที่นี้ผู้วิจัยได้ใช้วิธีการลบข้อมูลที่ขาดหายและข้อมูลที่รบกวนและค่าข้อมูลที่มีความผิดปกติออกจากฐานข้อมูลที่ใช้ในการวิเคราะห์ หลังจากผ่านขั้นตอนการเตรียมข้อมูลแล้วคงเหลือข้อมูลที่สมบูรณ์และสามารถนำไปสร้างตัวแบบหรือโมเดลได้จำนวน 3,000 ชุด

ตารางที่ 1 จำนวนข้อมูลที่นำมาใช้ในการวิจัย

ที่	ชื่อคุณลักษณะ	คำอธิบาย
1	Sex	เพศ
2	Age	อายุ
3	Education	ระดับการศึกษา
4	Major	สาขาวิชา
5	GPA	เกรดเฉลี่ยสะสม
6	Target	เป้าหมายทางการศึกษา
7	Address	ภูมิลำเนาเดิม
8	Father	อาชีพบิดา
9	Mdiploma	อาชีพมารดา
10	Avg income	รายได้เฉลี่ยต่อเดือน
11	Result	กลุ่มบุคลิกภาพ

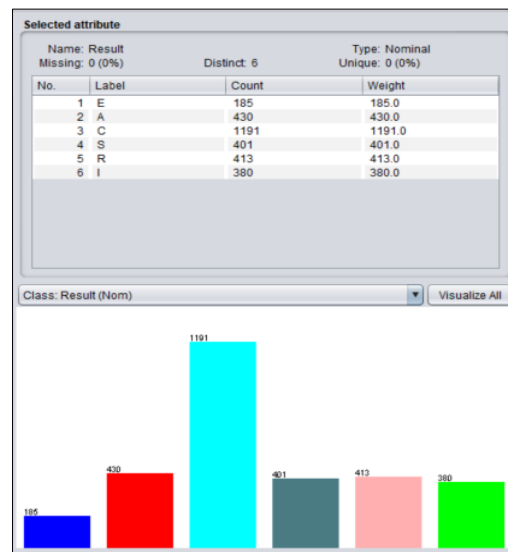
2.2.2.3 การคัดเลือกข้อมูล เป็นการคัดเลือกข้อมูลที่เหมาะสมตรงกับวัตถุประสงค์ของ

งานวิจัย เพื่อจะนำไปวิเคราะห์และทำการจัดกลุ่มข้อมูลออกเป็น 6 กลุ่มบุคลิกภาพผู้เรียนตามทฤษฎีของจอห์น ฮอลแลนด์ ทั้ง 6 กลุ่ม ได้แก่ กลุ่มบุคลิกภาพแบบจริงจัง (REALISTIC) กลุ่มบุคลิกภาพแบบที่ต้องใช้หาข้อเท็จจริง (INVESTIGATIVE) กลุ่มบุคลิกภาพแบบมีศิลปะ (ARTISTIC) กลุ่มบุคลิกภาพที่ชอบสมาคมและสังคม (SOCIAL) กลุ่มบุคลิกภาพแบบกล้าคิดกล้าทำ (ENTERPRISING) และกลุ่มบุคลิกภาพแบบยึดมั่นและมีระเบียบแบบแผน (Conventional)

2.2.2.4 การแปลงข้อมูล เป็นการแปลงข้อมูลที่เลือกมาให้อยู่ในรูปแบบที่สามารถนำไปใช้ในการวิเคราะห์ตามวิธีการทางด้านเหมืองข้อมูลได้ โดยได้กำหนดข้อมูลพื้นฐานของผู้เรียนซึ่งเป็นตัวแปรที่จะนำไปใช้ในการวิเคราะห์ด้วยเทคนิคเหมืองข้อมูล 11 ตัวแปร และมีการจัดเก็บข้อมูลของผู้เรียนไว้ในรูปแบบของไฟล์ *.CSV (Comma-Separated Value) ก่อนที่จะนำไปประมวลผลข้อมูลโดยใช้โปรแกรม WEKA ซึ่งเป็นโปรแกรมที่มีความสามารถทางด้านการทำเหมืองข้อมูล โดยโปรแกรมจะนำข้อมูลมาทำการวิเคราะห์โดยใช้อัลกอริทึมที่มีอยู่ในชุดเครื่องมือ เพื่อดึงเอาความรู้หรือตัวแปรที่สำคัญมาใช้ในการวิเคราะห์หรือทำนายสิ่งต่าง ๆ ซึ่งสามารถแสดงข้อมูลเพื่อจำแนกบุคลิกภาพของผู้เรียนตามทฤษฎีของจอห์น ฮอลแลนด์ ดังรูปที่ 1

sex	age	education	major	GPA	target	address	father	midploma	avg income	Result
F	low	voc	ma	D	Bachelor	Central	Government	Government	pay2	E
F	low	voc	ma	D	Bachelor	Central	Government	Government	pay2	A
M	low	voc	ac	B	diploma	BKK	Owner	Government	pay2	A
F	low	voc	ac	C	diploma	BKK	Owner	Government	pay2	A
F	low	voc	ac	C	diploma	South	Government	Government	pay1	A
F	low	voc	ac	D	diploma	Northeast	Company	Company	pay1	C
F	low	voc	ac	C	diploma	BKK	Company	Company	pay1	S
F	low	voc	ac	A	Bachelor	BKK	Government	Government	pay1	A
F	low	voc	ac	A	Bachelor	BKK	Government	Government	pay1	R
F	low	voc	ac	D	Vocational	South	SE	Government	pay2	A
M	low	voc	ac	A	Bachelor	Northeast	Company	Company	pay2	R
F	low	voc	ac	B	Bachelor	Northeast	Company	Company	pay2	C
F	low	voc	ma	B	Master	Northeast	Government	Company	pay2	R
F	low	voc	ac	B	diploma	BKK	Company	Company	pay2	A

รูปที่ 1 ตัวอย่างข้อมูลเพื่อจำแนกบุคลิกภาพของผู้เรียนตามทฤษฎีของจอห์น ฮอลแลนด์



รูปที่ 2 ผลการประเมินความถูกต้องของตัวจำแนกข้อมูลตามบุคลิกภาพต่างๆ

จากรูปที่ 2 ผู้วิจัยจะนำข้อมูลมาวิเคราะห์จากข้อมูลที่ได้จัดเก็บจำนวน 3,000 รายการ สามารถนำไปพัฒนาแบบจำลองได้ โดยสามารถแยกเป็นผู้เรียนที่มีกลุ่มบุคลิกภาพแบบจริงจัง เท่ากับ 413 รายการ กลุ่มบุคลิกภาพแบบที่ต้องใช้เซวปัญญา เท่ากับ 380 รายการ กลุ่มบุคลิกภาพแบบมีศิลปะ เท่ากับ 430 รายการ กลุ่มบุคลิกภาพที่ชอบสมาคมและสังคม เท่ากับ 401 รายการ กลุ่มบุคลิกภาพแบบกล้าคิดกล้าทำ เท่ากับ 185 รายการ และกลุ่มบุคลิกภาพแบบยึดมั่นและมีระเบียบแบบแผน เท่ากับ 1,191 รายการ

2.2.3 การสร้างแบบจำลอง

ในขั้นตอนนี้เป็น การนำเอาตัวแปรที่ได้ทำการทดสอบกับข้อมูลชุดศึกษา (Training Set) โดยใช้อัลกอริทึมในการจำแนกข้อมูล 5 อัลกอริทึม ดังนี้ เทคนิควิธีต้นไม้ตัดสินใจ ด้วยอัลกอริทึม J48 LMT และ Random Forest เทคนิคการเรียนรู้แบบเบย์ และเทคนิคโครงข่ายประสาทเทียม เหตุผลในการเลือกใช้อัลกอริทึมทั้ง 5 อัลกอริทึม เพราะเป็นเทคนิคเหมือนข้อมูลที่นิยมใช้กันมากในการจำแนกข้อมูล (Classification) เนื่องจากใช้งานง่าย มีความรวดเร็วในการจำแนกข้อมูล เหมาะสำหรับชุดข้อมูลที่มีความซับซ้อน และในการวิเคราะห์ข้อมูลยังมีความถูกต้องสูง [1] และในงานวิจัยนี้นำผลลัพธ์อัลกอริทึมทั้ง 5 มาเปรียบเทียบค่าความถูกต้องว่าเทคนิควิธีใดเหมาะสำหรับการวิเคราะห์รูปแบบความสนใจเลือกอาชีพของผู้เรียน โดยทำการทดสอบไขว้จำนวน 10 รอบ (10-fold cross-Validation) โดยใช้เกณฑ์ในการประเมินผลทดลองครั้งนี้ จำนวน 4 วิธี ได้แก่ การวัดค่าความถูกต้อง (Accuracy), ค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-measure) เพื่อนำไปใช้ในการจำแนกบุคลิกภาพของผู้เรียนต่อไป

ซึ่งเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจเป็นวิธีการเรียนรู้ของเครื่องมือที่นิยมใช้มากที่สุดแบบหนึ่ง [2] โดยการจำแนก (classification) ข้อมูลออกเป็นคลาส (class) ต่าง ๆ โดยใช้คุณสมบัติ (attribute) ของข้อมูลในการจำแนกว่าคุณสมบัติใดของข้อมูลที่เป็นตัวกำหนดการจำแนกและคุณสมบัติแต่ละตัวของข้อมูล

สำหรับเทคนิคการจำแนกข้อมูลด้วยวิธีการเรียนรู้แบบเบย์ เป็นเทคนิคที่ใช้ทฤษฎีความน่าจะเป็นตามกฎของเบย์ (Bayes' Theorem) เพื่อหาว่าสมมติฐานใดน่าจะถูกต้องที่สุด โดยใช้ความรู้ก่อนหน้า (Prior Knowledge) ได้แก่ ความน่าจะเป็นก่อนหน้าสำหรับสมมติฐานหนึ่ง ๆ ร่วมกับข้อมูล เช่น ความน่าจะเป็นที่สังเกตได้สำหรับสมมติฐานหนึ่ง ๆ เพื่อหาสมมติฐานที่ดีที่สุด

สำหรับโครงข่ายประสาทเทียม (Artificial Neural Networks) เป็นหนึ่งในเทคนิคของการทำเหมืองข้อมูล (Data Mining) คือ โมเดลทางคณิตศาสตร์ สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนกชันนิสต์ (Connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ ด้วยวัตถุประสงค์ที่จะสร้างเครื่องมือซึ่งมีความสามารถในการเรียนรู้การจดจำรูปแบบ (Pattern Recognition) และการสร้างความรู้ใหม่ (Knowledge Extraction) เช่นเดียวกับความสามารถที่มีในสมองมนุษย์

2.2.4 การวัดประสิทธิภาพของแบบจำลอง

ทำการทดสอบประสิทธิภาพในการจำแนกข้อมูลโดยการวัดความถูกต้องของทั้ง 5 อัลกอริทึม ด้วยวิธีการ 10-fold cross-Validation ซึ่งจะแบ่งข้อมูลออกเป็น 10 ชุดเท่ากัน ๆ กัน แล้วให้ใช้ 1 กลุ่มมาเป็นกลุ่มทดสอบ (test Set) ส่วนที่เหลือ 9 ชุด นำมาใช้เป็นกลุ่มเรียนรู้ (Training Set) แล้วทำ

การวนทำเป็นจำนวน 10 รอบ โดยเปลี่ยนกลุ่มทดสอบไปเรื่อย ๆ จนครบ การวัดประสิทธิภาพของการจำแนกประเภทข้อมูล โดยใช้เกณฑ์ในการประเมินผลทดลองครั้งนี้ จำนวน 4 วิธี ได้แก่ การวัดค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-measure)

การวิเคราะห์ประสิทธิภาพ คือ การวัดประสิทธิภาพการทำงานในแต่ละขั้นตอนวิธีสามารถวัดได้จากผลของการจำแนกกลุ่มข้อมูล โดยค่าของผลลัพธ์ที่ได้จากการจำแนกคือ ค่า True Positive (TP) ค่า True Negative (TN) ค่า False Positive (FP) ค่า False Negative (FN) และสามารถหาว่า

$$\text{ค่าความถูกต้อง (Accuracy) จากสมการ}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{ค่าความแม่นยำ (Precision) จากสมการ}$$

$$Precision = \frac{TP}{TP + FP}$$

$$\text{ค่าความระลึก (Recall) จากสมการ}$$

$$Recall = \frac{TP}{TP + FN}$$

$$\text{ค่าความถ่วงดุล (F-measure) จากสมการ}$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

3. ผลการวิจัย

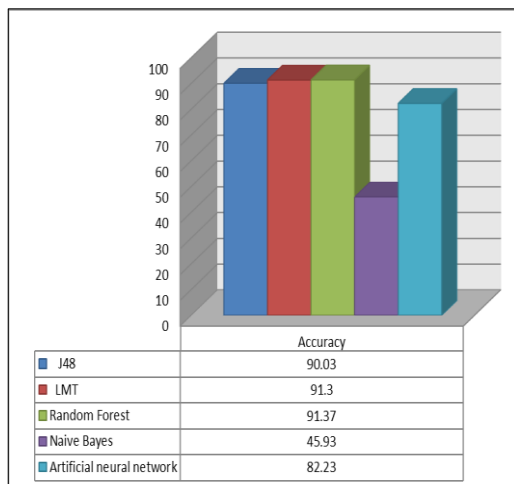
งานวิจัยนี้ทำการเปรียบเทียบอัลกอริทึมในการจำแนกข้อมูลที่มีประสิทธิภาพสูงสุดและทำการสร้างแบบจำลองที่มีผลต่อการเลือกอาชีพของผู้เรียน เพื่อเป็นแนวทางให้กับผู้เรียนในการใช้ประกอบการตัดสินใจ โดยใช้ข้อมูลนักศึกษาจากวิทยาลัยเทคโนโลยีบริหารธุรกิจจันทบุรี ทำการ

วิเคราะห์ด้วยเทคนิคเหมืองข้อมูล โดยใช้เทคนิคเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้อัลกอริทึม J48 LMT Random Forest เทคนิคการจำแนกข้อมูลด้วยวิธีการเรียนรู้แบบเบย์ และเทคนิคการจำแนกข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ด้วยวิธีการ 10-fold cross-Validation เพื่อหาค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล แล้วทำการเปรียบเทียบประสิทธิภาพของแต่ละเทคนิค

หลังจากทำการทดสอบแบบจำลองด้วยการวัดประสิทธิภาพจากผลการสร้างตัวแบบจำลองด้วยโปรแกรม WEKA จะทำให้ทราบค่าประสิทธิภาพต่าง ๆ ซึ่งสามารถนำตัวแบบจำลองที่ได้มาเปรียบเทียบเพื่อคัดเลือกตัวแบบจำลองที่เหมาะสม และมีความถูกต้องในการทำนายหรือพยากรณ์มากที่สุด รายละเอียดการทดสอบประสิทธิภาพสามารถแบ่งได้ 2 ส่วนคือ การเปรียบเทียบ ผลการวิเคราะห์ค่าความถูกต้องของแต่ละอัลกอริทึม และการเปรียบเทียบประสิทธิภาพของอัลกอริทึมสำหรับจำแนกข้อมูล การเปรียบเทียบผลการวิเคราะห์ค่าความถูกต้องของแต่ละเทคนิค โดยการวัดความถูกต้อง (Accuracy) ซึ่งสามารถสรุปผลค่าความถูกต้องได้ดังตารางที่ 1 และรูปที่ 1

ตารางที่ 2 ตารางเปรียบเทียบผลการวิเคราะห์ค่าความถูกต้องของแต่ละเทคนิค

Algorithms	Accuracy
Decision Tree	
J48	90.03
LMT	91.30
Random Forest	91.37
Naive Bayes	45.93
Artificial Neural Network	82.23



รูปที่ 3 แสดงกราฟการเปรียบเทียบผลการวิเคราะห์ค่าความถูกต้องของแต่ละเทคนิค

จากตารางที่ 2 แสดงผลการวิเคราะห์ค่าความถูกต้องของแต่ละเทคนิค ซึ่งเมื่อพิจารณาความถูกต้องของการจำแนกข้อมูล พบว่าประสิทธิภาพในการจำแนกข้อมูลด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ด้วยอัลกอริทึม Random Forest ให้ค่าความถูกต้องสูงที่สุดร้อยละ 91.37 และวิธีการเรียนรู้แบบเบย์ให้ค่าความถูกต้องน้อยที่สุดในการจำแนกข้อมูล

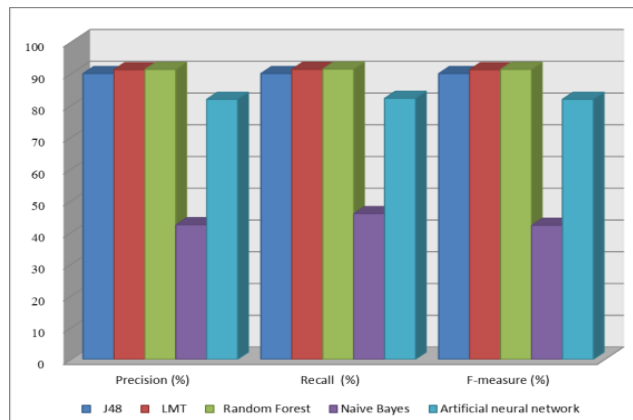
การเปรียบเทียบประสิทธิภาพของเทคนิคสำหรับจำแนกข้อมูล โดยการวัดค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-measure) ซึ่งสามารถสรุปผลได้ดังตารางที่ 3 และรูปที่ 4

จากตารางที่ 3 แสดงผลการเปรียบเทียบประสิทธิภาพของอัลกอริทึมสำหรับการจำแนกข้อมูลความสนใจอาชีพของผู้เรียนพบว่าความแม่นยำ (Precision) ของแบบจำลองด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ด้วยอัลกอริทึม Random Forest ให้ค่าความแม่นยำสูงที่สุดร้อยละ 91.30 และวิธีการเรียนรู้แบบเบย์ให้ค่าความแม่นยำน้อยที่สุดในของจำแนกข้อมูล

การวัดค่าความระลึก (Recall) พบว่าเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจด้วยอัลกอริทึม Random Forest ให้ค่าความระลึกสูงที่สุดร้อยละ 91.40 และวิธีการเรียนรู้แบบเบย์ให้ค่าความระลึก น้อยที่สุดในของจำแนกข้อมูล

ตารางที่ 3 การเปรียบเทียบประสิทธิภาพของอัลกอริทึมสำหรับจำแนกข้อมูล

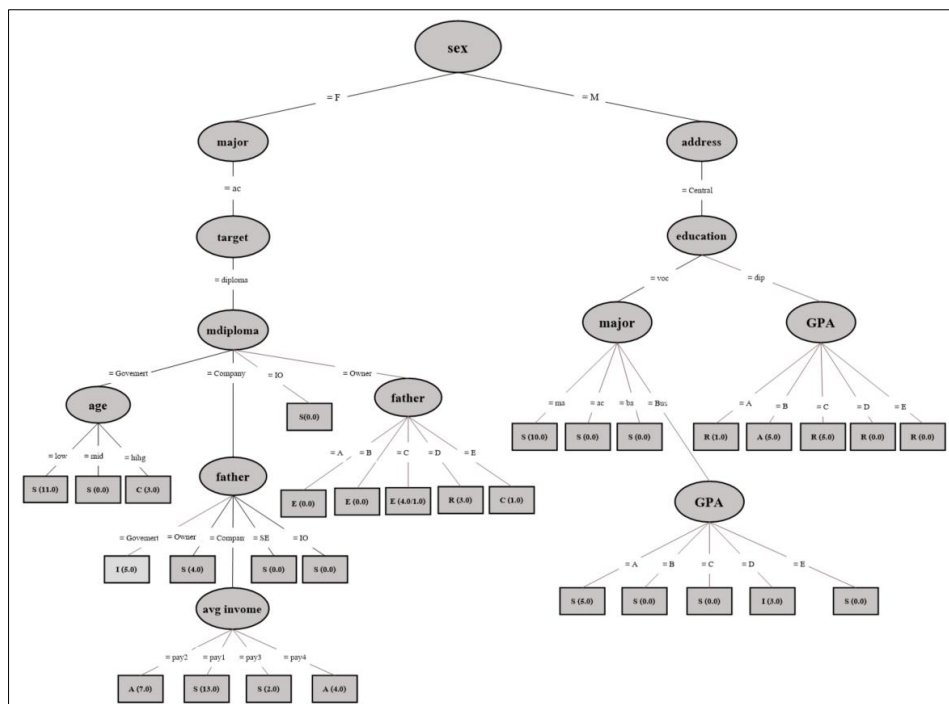
Algorithms	Precision (%)	Recall (%)	F-measure (%)
Decision Tree			
J48	90.00	90.00	90.00
LMT	91.20	91.30	91.20
Random Forest	91.30	91.40	91.30
Naive Bayes	42.40	45.90	42.20
Artificial neural network	81.90	82.20	81.90



รูปที่ 4 กราฟเปรียบเทียบประสิทธิภาพของอัลกอริทึมสำหรับจำแนกข้อมูล

จากการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลพบว่าเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ด้วยอัลกอริทึม Random Forest ให้ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล ในการจำแนกข้อมูลมีประสิทธิภาพสูงที่สุด ด้วยเหตุนี้ผู้วิจัยจึงนำเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ด้วยอัลกอริทึม

ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ด้วยอัลกอริทึม Random Forest มาใช้ในการวิจัยครั้งนี้ สามารถแสดงตัวอย่างของแบบจำลองต้นไม้ตัดสินใจด้วยอัลกอริทึม Random Forest ดังรูปที่ 5 ซึ่งโครงสร้างต้นไม้ตัดสินใจนี้จะนำไปใช้ในการพัฒนาระบบการแนะนำอาชีพของผู้เรียนต่อไป



รูปที่ 5 ตัวอย่างของแบบจำลองต้นไม้ตัดสินใจด้วยอัลกอริทึม Random Forest

4. สรุปผลการวิจัย

งานวิจัยนี้ได้ทำการวิเคราะห์รูปแบบความสนใจเลือกอาชีพของผู้เรียน โดยใช้แบบทดสอบตามทฤษฎีของจอห์น ฮอลแลนด์โดยการประยุกต์ใช้เทคนิคเหมืองข้อมูลเพื่อหาตัวแบบจำลองที่เหมาะสมที่สุด ซึ่งทำการเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลความสนใจเลือกอาชีพของผู้เรียน ด้วยวิธีการจำแนกข้อมูล โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจเลือกอัลกอริทึม J48 LMT Random Forest เทคนิคการจำแนกข้อมูลด้วยวิธีการเรียนรู้แบบเบย์ และเทคนิคการจำแนกข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ซึ่งตัวแปรที่ใช้ในการจำแนกข้อมูลครั้งนี้มี 11 ตัวแปร ได้แก่ เพศ อายุ ระดับการศึกษา สาขาวิชา เกรดเฉลี่ยสะสม เป้าหมายทางการศึกษา ภูมิถิ่นอาศัย อาชีพบิดา อาชีพมารดา รายได้เฉลี่ยต่อเดือนของครอบครัว และกลุ่มบุคลิกภาพ ข้อมูลที่ใช้ในการจำแนก ได้จากการเก็บรวบรวมข้อมูลของนักศึกษาวิทยาลัยเทคโนโลยีบริหารธุรกิจจันทบุรี จากผลการทดลองทั้งหมดพบว่า เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ด้วยอัลกอริทึม Random Forest ให้ประสิทธิภาพในการจำแนกข้อมูลได้ดีกว่าทุกเทคนิคที่เปรียบเทียบ เนื่องจากเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ เป็นเทคนิคหนึ่งในการเรียนรู้ของเครื่อง ซึ่งอาศัยการแตกเงื่อนไขย่อยไปเรื่อย ๆ เพื่อนำไปสู่คำตอบ ซึ่งส่งผลให้ตัวจำแนกสามารถจำแนกข้อมูลที่มีคุณลักษณะที่แตกต่างกัน และส่งผลทำให้ตัวจำแนกให้ประสิทธิภาพในการจำแนกข้อมูลสูงกว่าอัลกอริทึมอื่น ส่งผลให้ประสิทธิภาพในการวัดค่าความถูกต้องสูงที่สุดเท่ากับ 91.37% ค่าความแม่นยำเท่ากับ 0.913 ค่าความระลึกลับเท่ากับ 0.914 และค่าความถ่วงดุลเท่ากับ

0.913 ได้ค่าสูงกว่าอัลกอริทึมอื่นที่ทำการเปรียบเทียบ จากผลการวิจัยสามารถสรุปได้ว่าตัวแบบเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ด้วยอัลกอริทึม Random Forest เป็นตัวแบบที่เหมาะสมที่จะนำไปพัฒนาเป็นระบบการแนะนำอาชีพของผู้เรียนต่อไป

อย่างไรก็ตาม งานวิจัยนี้เสนอการจำแนกข้อมูลด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ด้วยอัลกอริทึม Random Forest ในการจำแนกข้อมูลด้วยตัวอย่างในการจำแนกไม่มากนัก หากสามารถนำไปทดสอบกับชุดข้อมูลตัวอย่างที่มีขนาดใหญ่ขึ้น จะสามารถแสดงให้เห็นถึงประสิทธิภาพในการจำแนกที่ชัดเจนมากขึ้น

5. เอกสารอ้างอิง

- [1] รัชดา ประดับศรี. การสร้างแบบประเมินสมรรถนะในการแนะแนวของเจ้าหน้าที่แนะแนวอาชีพ. ปรินธิพนธ์ การศึกษามหาบัณฑิต สาขาจิตวิทยาการแนะแนว มหาวิทยาลัยศรีนครินทรวิโรฒ. 2552.
- [2] พัฒนพงษ์ คลรรัตน์.จารี ทองคำ. การเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์ความสำเร็จการศึกษาของนักเรียนระดับประกาศนียบัตรวิชาชีพ. วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม.37(3): 378-388. 2018.
- [3] กรมส่งเสริมการค้าส่งออก กระทรวงพาณิชย์. คู่มือการค้าและการลงทุนในประเทศอาเซียน. กรมส่งเสริมการค้าส่งออก กระทรวงพาณิชย์.2555.

- [4] Ansari, Gufran Ahmad. Career Guidance through Multilevel Expert System Using Data Mining Technique. International Journal of Information Technology and Computer Science. 2017. 9(8): 22-29.
- [5] Nikita Gorad. Ishani Zalte. Aishwarya Nandi. & Deepali Nayak. Career Counselling Using Data Mining. International Journal of Innovative Research in Computer and Communication Engineering. 5(4). 2017.
- [6] Holland, John L. 1966. The Psychology of Vocational Choice: A Theory of Personality Types and Model Environments. Waltham, Mass: Blaisdell, 1966.
- [7] Holland John L. Making Vocational Choice: A Theory of Careers. Englewood Cliffs, NJ: Prentice-Hall. 1973.
- [8] ชัชชฎา วันดี. การศึกษาปัจจัยที่มีผลต่อการเลือกอาชีพของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษา โด ช ใ ช เท ค นิก เหมืองข้อมูล. วิทยานิพนธ์บัณฑิต มหาวิทยาลัยมหาสารคาม. 2557.
- [9] สำราญ วานนท์ ธวัช อวีร์ราษฎร์ และจรัญ แสนราช. การศึกษาเทคนิคพยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีสาขาคอมพิวเตอร์โดยใช้เทคนิคเหมืองข้อมูล. วารสารวิชาการการจัดการเทคโนโลยีสารสนเทศและนวัตกรรม. 2561. 5(1) :163-171.