

การเปรียบเทียบประสิทธิภาพโมเดลในการเลือกสาขาการเรียนต่อระดับ
ประกาศนียบัตรวิชาชีพ โดยใช้เทคนิคการจำแนกข้อมูล
Comparison Performance of Selected Field of Study Model for
Vocation Institution using Data Classification

ชนะวงศ์ คงสอน (Chanawong Kongsorn), ณัฐวุฒิ วิสุทธิพิเนตร (Nattawut Visutti-pinat)

และองอาจ อุ่นอนันต์ (Aongart Aun-a-nan)

สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ

fream_bam@hotmail.com, paoonline@windowslive.com, aongart.a@rmutsb.ac.th

บทคัดย่อ

การเลือกสาขาเรียนต่อของนักศึกษาอาชีวะ จะส่งผลโดยตรงต่อผลสัมฤทธิ์ทางการเรียนของนักศึกษาเองถ้าเลือกเรียนในสาขาที่ตรงกับความถนัด ซึ่งยังขาดเครื่องมือที่จะช่วยให้ผู้สมัครที่ยังเลือกสาขาเรียนต่อไม่ได้ ผู้วิจัยจึงพัฒนาระบบสนับสนุนการตัดสินใจเลือกสาขาเรียน แต่จะต้องทำการเปรียบเทียบประสิทธิภาพของโมเดลการจำแนกข้อมูล ทั้ง 3 ได้แก่ K-nearest Neighbors, Decision Tree และ Rule Based ซึ่งผลการเปรียบเทียบประสิทธิภาพโมเดลที่ได้จากเทคนิค K-nearest Neighbors (K=6) มีความถูกต้องมากที่สุด 76.62% ซึ่งจะนำไปพัฒนาระบบสนับสนุนการตัดสินใจต่อไปในอนาคต

ABSTRACT

The selection of study of vocational students will directly affect the achievement of students choose to study in the field if matching aptitude. Lack of tools to help students for field not selected. Researchers have developed a decision support system for the field of study. But compare the performance of the three data classification models consist of K-nearest Neighbors, Decision Tree and Rule Based. The results showed K-nearest Neighbors (k=6) got the high performance have accuracy score 76.62%. This can lead to the development of decision support systems in the future.

คำสำคัญ— เหมือนข้อมูล; การจำแนกข้อมูล; ประสิทธิภาพ; ทฤษฎีการเลือกอาชีพ

1. บทนำ

ในการเลือกสาขาการเรียนของผู้ที่ต้องการสมัครเข้าศึกษาต่อในระดับประกาศนียบัตรวิชาชีพ (ปว.ช.) ในระบบอาชีวศึกษา นับเป็นสิ่งที่สำคัญอย่างมากเนื่องจากจะส่งผลโดยตรงต่อผู้สมัครเอง ซึ่งในกรณีที่ตัดสินใจเลือกเรียนในสาขาที่ไม่เหมาะสมกับผู้สมัครเองมีหลายสาเหตุ เช่น ผู้สมัครบางรายไม่รู้ความถนัดหรือความชอบของตนเองทำให้เลือกสาขาการเรียนตามเพื่อน ผู้สมัครบางรายทางครอบครัวอยากให้เรียนทางด้านที่ผู้สมัครเองไม่ชอบ เป็นต้น ซึ่งในบางรายส่งผลให้ไม่สัมฤทธิ์ผลทางการเรียนทำให้เป็นการเสียเวลาในการเรียนหรือจบการศึกษาไปอย่างด้อยคุณภาพ ดังนั้นการเลือกสาขาเรียนตามความสามารถ ความถนัดหรือความสนใจของตนเองอย่างแท้จริงจะทำให้มีโอกาสสัมฤทธิ์ผลทางการเรียนมากขึ้น จากปัญหาที่กล่าวมาประเด็นสำคัญคือทำอย่างไรจึงสามารถวิเคราะห์ความถนัดที่อยู่ภายในตัวของผู้สมัครเองออกมาและทำการแนะนำให้ผู้สมัครรู้ความสามารถของตนเองว่าเหมาะสมกับสาขาใดมากที่สุด

จากการศึกษางานวิจัยต่างๆพบว่าการทำเหมือนข้อมูลเป็นเทคนิคที่นิยมใช้ในการวิเคราะห์และสกัดหาองค์ความรู้จากข้อมูลเดิมที่มีอยู่ [1] โดยการทำเหมือนข้อมูลนี้มีองค์ประกอบและทฤษฎีต่างๆที่สามารถนำมาใช้งานได้อย่างหลากหลาย ซึ่งการนำเทคนิคการจำแนกข้อมูล (Data Classification) มาใช้ในการวิเคราะห์เพื่อทำการจำแนกข้อมูลต่างๆ [2] มีการนำไป

ประยุกต์ใช้งานได้หลากหลาย เช่น การพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนักศึกษาคณะวิศวกรรมศาสตร์ [3] หรือการจำแนกเนื้อหาอีเมลที่เป็นสแปมเมลล์ ซึ่งสถาบันอาชีวศึกษาในจังหวัดพระนครศรีอยุธยา ยังไม่มีระบบที่ช่วยให้ผู้สมัครตัดสินใจเลือกสาขาการเรียนได้ตรงตามความถนัดของตนเอง

ในงานวิจัยนี้ได้นำเสนอการเปรียบเทียบประสิทธิภาพโมเดลในการเลือกสาขาการเรียนต่อระดับประกาศนียบัตรวิชาชีพ โดยใช้เทคนิคการจำแนกข้อมูล เพื่อนำโมเดลที่มีประสิทธิภาพที่ดีที่สุดไปพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจเลือกสาขาการเรียน โดยใช้ข้อมูลของนักศึกษาระดับประกาศนียบัตรวิชาชีพของแต่ละสาขา สังกัดอาชีวศึกษาจังหวัดพระนครศรีอยุธยา เพื่อวิเคราะห์และแนะนำให้ผู้สมัครรู้ความสามารถของตนเองว่าเหมาะสมกับสาขาใดมากที่สุด

ในบทความวิจัยนี้แบ่งเนื้อหาออกเป็น ดังนี้ ส่วนที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง ส่วนที่ 3 วิธีการดำเนินงาน ส่วนที่ 4 ผลการดำเนินงาน และส่วนที่ 5 สรุปผลงานวิจัยและข้อเสนอแนะ

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

จากการศึกษาการเปรียบเทียบประสิทธิภาพโมเดลในการเลือกสาขาการเรียนต่อระดับประกาศนียบัตรวิชาชีพ โดยใช้เทคนิคการจำแนกข้อมูล มีทฤษฎีที่เกี่ยวข้อง ดังนี้

2.1. การสัมฤทธิ์ผลทางการเรียนของนักศึกษา

การสัมฤทธิ์ผลทางการเรียน หมายถึง การศึกษาครบจำนวนหน่วยกิตตามหลักสูตร และได้รับอนุมัติปริญญาบัตรจากสภามหาวิทยาลัย โดยมี 2 ลักษณะ คือ สำเร็จการศึกษาตามเวลาที่หลักสูตรกำหนด และสำเร็จการศึกษาโดยใช้เวลามากกว่าที่หลักสูตรกำหนด [4]

รายวิชาเป็นตัวชี้วัดหนึ่งในการสัมฤทธิ์ทางการเรียน การเรียนวิชาที่ตนเองมีความถนัดนั้นจะช่วยให้เกรดเฉลี่ยไปในทางที่ดีแล้วยังช่วยให้ตนเองค้นพบว่าชอบหรือมีความถนัดไปทางวิชาไหน ในกรณีที่ต้องการศึกษาต่อก็จะทำให้ตัดสินใจในการเลือกสาขาการเรียนต่อได้ไม่ยาก [5]

2.2. ทฤษฎีการเลือกอาชีพของ John Holland

เป็นผลจากการสังเกตของเขาและของคนอื่น ๆ เกี่ยวกับความสนใจ ลักษณะและพฤติกรรมของบุคคล และมีส่วนสัมพันธ์กับทฤษฎีของกิลฟอร์ด (Guilford) [6] ซึ่งได้วิเคราะห์บุคลิกภาพและความสนใจของบุคคลออกเป็น 6 ประเภท [7] ดังนี้

2.2.1. บุคลิกภาพแบบนิยมความจริง (Realistic: R) ลักษณะโดยทั่วไป ชอบทำกิจกรรมที่ต้องใช้พลังกำลัง ชอบทำงานกลางแจ้ง ชอบกิจกรรมที่ไม่สลับซับซ้อน สาขาวิชาที่เหมาะสม ได้แก่ สาขาวิชาด้านช่าง เครื่องยนต์ การเกษตร ไฟฟ้า และเทคนิค

2.2.2. บุคลิกภาพแบบช่างคิด (Investigative: I) ลักษณะโดยทั่วไป ชอบคิด สังเกต วิเคราะห์ ชอบคิดปัญหาวิจารณ์อย่างมีเหตุผล ชอบใฝ่หาความรู้ สาขาวิชาที่เหมาะสม ได้แก่ สาขาวิชาด้านวิทยาศาสตร์ การคำนวณ คอมพิวเตอร์

2.2.3. บุคลิกภาพแบบรักศิลปะ (Artistic: A) ลักษณะโดยทั่วไป ชอบกิจกรรมเกี่ยวกับนามธรรม เป็นอิสระ รักความงาม มีความเป็นตัวของตัวเองชอบใช้ชีวิตและกิจกรรมตามลำพัง สาขาวิชาที่เหมาะสม ได้แก่ สาขาวิชาด้านศิลปะ ภาษา ดนตรี การละคร การเขียน

2.2.4. บุคลิกภาพแบบชอบเข้าสังคม (Social: S) ลักษณะโดยทั่วไป ชอบติดต่อกับผู้อื่น ชอบสนทนา ชอบให้ความรู้ สาขาวิชาที่เหมาะสม ได้แก่ สาขาวิชาด้านภาษา มนุษยสัมพันธ์ งานประชาสัมพันธ์ งานบริการวิชาการ

2.2.5. บุคลิกภาพแบบกล้าเสี่ยง (Enterprising: E) ลักษณะโดยทั่วไป มีความเป็นผู้นำ มีความคิดริเริ่ม กล้าคิด กล้าทำ มีทักษะในการพูด ชอบงานสังคม สาขาวิชาที่เหมาะสม ได้แก่ สาขาวิชาด้านการตลาด ธุรกิจ เศรษฐกิจและการเมือง

2.2.6. บุคลิกภาพแบบจารีตนิยม (Conventional: C) ลักษณะโดยทั่วไป ชอบใช้กิจกรรมเป็นรูปธรรมและกิจกรรมทางภาษา ยึดระเบียบ ชอบทำตามระเบียบแบบแผนมากกว่าการริเริ่มด้วยตนเอง เป็นพวกวัตถุนิยมและเจ้าระเบียบ สาขาวิชาที่เหมาะสม ได้แก่ สาขาวิชาด้านงานสารบรรณ การคำนวณ งานบัญชี

2.3. การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล คือ ขบวนการทำงานในการกลั่นกรองข้อมูลจากฐานข้อมูลขนาดใหญ่ (Large Information) เพื่อให้ได้

สารสนเทศที่ยังไม่รู้ เป็นสารสนเทศที่สามารถนำไปใช้ประโยชน์ได้ (Actionable) ซึ่งเป็นสิ่งสำคัญที่ช่วยตัดสินใจในการทำธุรกิจ โดยการทำเหมืองข้อมูลเป็นขั้นตอนที่สำคัญในการค้นหาคำตอบ ความรู้ในฐานข้อมูล (Knowledge Discovery in Database: KDD) [2] เช่น การวิเคราะห์พฤติกรรมในการซื้อสินค้าของผู้บริโภค การพยากรณ์อากาศ เป็นต้น

2.4. การจำแนกข้อมูล (Classification)

การทำเหมืองข้อมูลเป็นการสกัดเอาสิ่งที่มีประโยชน์ออกมาจากข้อมูลที่มีจำนวนมาก ซึ่งมีหลากหลายวิธีหนึ่งในนั้นคือการจำแนกข้อมูล (Classification) ซึ่งมีเทคนิคต่าง ๆ ที่นิยมใช้งาน เช่น Decision Tree, Naive Bayes, K-Nearest Neighbors และ Neural Network โดยขั้นตอนแรกในการจำแนกข้อมูลให้นำข้อมูลที่ใช้ในการเรียนรู้ (Training Data) มาสร้างเป็นโมเดลขึ้นมาด้วยเทคนิคการจำแนกข้อมูลแบบต่างๆ และทำการวัดประสิทธิภาพของโมเดล ดังนี้

2.4.1. Confusion Matrix คือ ตารางแบบจัตุรัสโดยมีจำนวนแถวเท่ากับจำนวนคอลัมน์และเท่ากับจำนวนคลาส [8] เช่น มีคลาสคำตอบอยู่ 2 คำ คือ Up และ Down สามารถสร้าง Confusion Matrix ได้ แสดงดังตารางที่ 1

ตารางที่ 1. แสดงตาราง Confusion Matrix

Predicted / Actual	Up	Down
Up	TP	FP
Down	FN	TN

ซึ่งประกอบไปด้วย ค่าจำนวนข้อมูลที่ทำนายถูกว่าเป็นคลาสซึ่งกำลังสนใจอยู่ (True Positive: TP), ค่าจำนวนข้อมูลที่ทำนายถูกว่าเป็นคลาสซึ่งไม่ได้สนใจอยู่ (True Negative: TN), ค่าจำนวนข้อมูลที่ทำนายผิดมาเป็นคลาสซึ่งกำลังสนใจอยู่ (False Positive: FP) และค่าจำนวนข้อมูลที่ทำนายผิดมาเป็นคลาสซึ่งไม่ได้สนใจอยู่ (False Negative: FN) [8]

2.4.2. ตัววัดประสิทธิภาพของโมเดลการจำแนกข้อมูล การนำโมเดลไปใช้งานจริงนั้นจะต้องทำการวัดประสิทธิภาพของโมเดลเสียก่อนโดยทั่วไปแล้วจะมีตัววัดที่นิยมใช้กันในงานวิจัยและการทำงานต่างๆ อยู่ 3 คำ [8] คือ

ค่าความถูกต้อง (Accuracy) เป็นการวัดความถูกต้องของโมเดล โดยพิจารณารวมทุกคลาส แสดงดังสมการ 1

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

ค่าความแม่นยำ (Precision) เป็นการวัดความแม่นยำของโมเดล โดยพิจารณาแยกทีละคลาส แสดงดังสมการ 2

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

ค่าค้นคืน (Recall) เป็นการวัดการค้นคืนของโมเดล โดยพิจารณาแยกทีละคลาส แสดงดังสมการ 3

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

2.5. งานวิจัยที่เกี่ยวข้อง

ไพฑูรย์ [2] ได้พัฒนาระบบสนับสนุนการตัดสินใจในการเลือกสาขาการเรียน ระดับปริญญาตรี โดยใช้เทคนิคต้นไม้ตัดสินใจ ซึ่งจากการทดลอง พบว่าในการสร้างตัวแบบควรแยกสร้างตัวแบบสำหรับแต่ละสาขาการเรียนเนื่องจากคุณสมบัติของผู้เรียนแต่ละสาขามีความแตกต่างกัน เพื่อให้ได้ตัวแบบที่สามารถทำนายแนวโน้มของผลการเรียนที่เหมาะสมสำหรับแต่ละสาขา

ชาลี [1] ได้พัฒนาระบบสนับสนุนการตัดสินใจเลือกประกอบอาชีพสำหรับผู้สำเร็จการศึกษาระดับปริญญาตรีด้วยวิธีการเรียนรู้แบบเบย์ โดยคำนวณหาความน่าจะเป็นที่จะเลือกประกอบอาชีพต่างๆ ข้อมูลที่นำมาใช้ในการเรียนรู้ได้มาจากการสำรวจข้อมูลของสำนักวิจัยและบริการวิชาการ วิธีการเรียนรู้แบบเบย์พบว่าตัวจำแนกประเภทเบย์อย่างง่าย สามารถนำมาใช้ในการทำนายการเลือกอาชีพได้เป็นอย่างดี

เรือนทิพย์ [9] ได้พัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบวิธีต้นไม้การตัดสินใจสำหรับการเลือกสาขาวิชาเรียนของนักศึกษาระดับอุดมศึกษา ในงานนี้มีความมุ่งหมายเพื่อสร้างตัวแบบสำหรับหาปัจจัยที่มีผลต่อการตัดสินใจในการเลือกสาขาวิชาเรียนของนักศึกษา และการพัฒนาตัวแบบภายใต้เทคนิคการจัดทำเหมืองข้อมูล ซึ่งในการทดสอบแบบจำลองที่ได้จะทำการ

ทดสอบผลพื้นฐานของวิธี C4.5(J48) โดยใช้โปรแกรม WEKA ในการสร้างแบบจำลอง หลังจากนั้นได้นำผลลัพธ์ที่ดีที่สุดจากเทคนิคต้นไม่การตัดสินใจมาสร้างตัวแบบ

3. วิธีการดำเนินงาน

กระบวนการนี้จะอธิบายถึงวิธีการดำเนินงาน โดยปริยญาณิพนธ์นี้ได้นำเทคนิคเหมืองข้อมูลมาใช้ในการพัฒนาระบบสนับสนุนการตัดสินใจเลือกสาขาการเรียนต่อ ระดับประกาศนียบัตรวิชาชีพ ซึ่งข้อมูลที่ใช้อยู่คือข้อมูลที่ได้จากการ ประเมินความถนัดรายวิชา และแบบทดสอบเลือกสาขาการเรียนของนักศึกษาในระดับประกาศนียบัตรวิชาชีพ

3.1. ศึกษาและรวบรวมทฤษฎีที่เกี่ยวข้อง

ทางผู้วิจัยได้ทำการศึกษาเกี่ยวกับเทคนิคเหมืองข้อมูล และทฤษฎีการเลือกอาชีพของ John Holland แล้วทำการแบ่งกลุ่มเพื่อเก็บข้อมูลตามสาขาการเรียนให้เข้ากับ 6 กลุ่มตามทฤษฎีการเลือกอาชีพของ John Holland จำนวนกลุ่มละ 84 คน รวมทั้งหมด 525 คน แสดงดังตารางที่ 2

ตารางที่ 2. แสดงรายละเอียดการแบ่งกลุ่มตามทฤษฎีจอห์น ฮอลแลนด์

หมวด	สาขาการเรียน	ระดับชั้น	จำนวน
A	วิจิตรศิลป์	ปว.ช.	42
A	ช่างเขียนแบบเครื่องกล	ปว.ช.	42
C	บัญชี	ปว.ช.	42
C	คอมพิวเตอร์ธุรกิจ	ปว.ช.	42
E	การตลาด	ปว.ช.	84
R	ช่างยนต์	ปว.ช.	21
R	ช่างกลโรงงาน	ปว.ช.	21
R	ช่างเชื่อมโลหะ	ปว.ช.	21
R	ช่างก่อสร้าง	ปว.ช.	21
R	ช่างซ่อมบำรุงเครื่องจักรกล	ปว.ช.	21
I	ช่างไฟฟ้ากำลัง	ปว.ช.	28
I	ช่างอิเล็กทรอนิกส์	ปว.ช.	28
I	เทคโนโลยีสารสนเทศ	ปว.ช.	28
S	ภาษาต่างประเทศ	ปว.ช.	42
S	การโรงแรม	ปว.ช.	42
รวมทั้งสิ้น			525

3.2. พัฒนาแบบประเมินและแบบทดสอบ

การพัฒนาแบบประเมินและแบบทดสอบเชิงจิตวิทยานั้น ได้ใช้รายวิชาโดยอ้างอิงจากหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 [5] และทฤษฎีการเลือกอาชีพของจอห์น ฮอลแลนด์ ซึ่งได้รับคำปรึกษาและการประเมินจากผู้เชี่ยวชาญด้านจิตวิทยา

3.3. เก็บรวบรวมข้อมูล

นำแบบประเมินและแบบทดสอบเชิงจิตวิทยา ไปให้นักศึกษา ระดับประกาศนียบัตรวิชาชีพประเมินและนำข้อมูลการประเมินของนักศึกษามาเก็บไว้ในฐานข้อมูลกำหนดคุณสมบัติ (Attribute) ที่จะนำมาใช้ในการสร้างโมเดลโดยมีรายละเอียดแสดงดังตารางที่ 3

ตารางที่ 3. แสดงรายละเอียดการเก็บข้อมูลจากแบบประเมิน

Attribute	Description	Data type
No.	ลำดับ	Number
Gender	เพศ	Text
Grade	เกรดเฉลี่ย (ม.ต้น)	Number
Thai	ความชอบกลุ่มวิชาภาษาไทย	Number
Math	ความชอบกลุ่มวิชาคณิตศาสตร์	Number
Sci	ความชอบกลุ่มวิชาวิทยาศาสตร์	Number
Social	ความชอบกลุ่มวิชาสังคมศึกษา ศาสนา และวัฒนธรรม	Number
Health	ความชอบกลุ่มวิชาสุขศึกษาและพลศึกษา	Number
Art	ความชอบกลุ่มวิชาศิลปะ	Number
Technology	ความชอบกลุ่มวิชาการงานอาชีพและเทคโนโลยี	Number
Languages	ความชอบกลุ่มวิชาภาษาต่างประเทศ	Number
Score A	คะแนนบุคลิกภาพแบบรักศิลปะ	Number
Score C	คะแนนบุคลิกภาพแบบจารีตนิยม	Number
Score E	คะแนนบุคลิกภาพแบบกล้าเสี่ยง	Number
Score I	คะแนนบุคลิกภาพแบบช่างคิด	Number
Score R	คะแนนบุคลิกภาพแบบนิยมความจริง	Number
Score S	คะแนนบุคลิกภาพชอบเข้าสังคม	Number
Major	สาขาการเรียน	Text
Class	คลาสที่ใช้ทำนาย	Text

3.4. สร้างโมเดลการจำแนกข้อมูล

การสร้างโมเดลการจำแนกข้อมูลนั้นใช้เครื่องมือ RapidMiner ในการสร้างโมเดล จากนั้นทำการแบ่งข้อมูลจาก Data Set โดยแบ่งเป็น Training Set และ Test Set เพื่อสร้างโมเดล โดยใช้เทคนิคการแบ่งข้อมูลด้วยวิธี 10 Fold Cross-Validation ซึ่งวิธีนี้จะเหมาะกับข้อมูลที่ไม่มาก เพราะจะทำให้ทดสอบข้อมูลได้ครบถ้วน โดยทำการแบ่งข้อมูลออกเป็นกลุ่มและทำการวนรอบทดสอบที่ละส่วนไปเรื่อยๆ จนครบทั้งหมด 10 ครั้ง เมื่อเราทำการแบ่งข้อมูลแล้วโดยเลือกใช้เทคนิคในการสร้างทั้งหมด 3 เทคนิค ได้แก่ K-nearest Neighbors (KNN), Decision Tree (DT) และ Rule Based (RB) มาสร้างโมเดล

3.5. ทดสอบและเปรียบเทียบประสิทธิภาพ

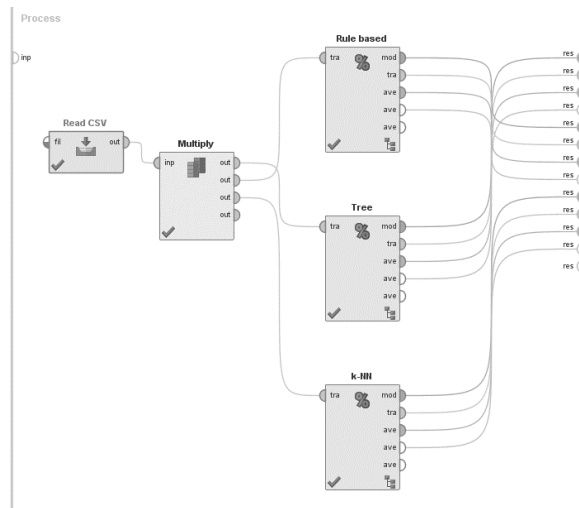
เมื่อสร้างโมเดลเสร็จแล้วทำการวัดค่าประสิทธิภาพโดยใช้ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และค่าเรียกค้นคืน (Recall) ซึ่งทำการปรับค่าพารามิเตอร์ต่างๆแล้วนำค่าที่ได้มาทำการเปรียบเทียบว่าโมเดลใดมีค่าประสิทธิภาพมากที่สุดจึงเลือกใช้โมเดลนั้นมาพัฒนาระบบตัดสินใจในการเลือกสาขาการเรียน

4. ผลการดำเนินงาน

ผู้วิจัยได้ดำเนินการเปรียบเทียบประสิทธิภาพโมเดลในการเลือกสาขาการเรียนต่อระดับประกาศนียบัตรวิชาชีพจากการ โดยทำการสร้างโมเดล ทำการปรับพารามิเตอร์ในแต่ละเทคนิคและบันทึกค่าประสิทธิภาพ มีรายละเอียดดังนี้

4.1. ผลการสร้างโมเดล

ในการสร้างโมเดลด้วย RapidMiner โดยใช้การแบ่งข้อมูลด้วยวิธี 10 Fold Cross-Validation โดยนำ 3 เทคนิคที่ใช้ประกอบไปด้วยเทคนิค K-nearest Neighbors, Decision Tree และ Rule Based แล้ววัดประสิทธิภาพด้วยค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และค่าเรียกค้นคืน (Recall) สามารถทำงานได้อย่างถูกต้อง แสดงดังรูปที่ 1



รูปที่ 1. แสดงการเชื่อมต่อระหว่าง Data Set กับ วิธี Cross-validation

4.2. ผลการหาค่าประสิทธิภาพเทคนิค K-nearest Neighbors

จากการทดลองหาค่าประสิทธิภาพโดยใช้เทคนิค K-nearest Neighbors ทำให้ได้ค่าประสิทธิภาพซึ่งประกอบไปด้วยค่าความถูกต้อง ค่าความแม่นยำ และค่าค้นคืน แสดงดังตารางที่ 4

ตารางที่ 4. แสดงเปรียบเทียบค่าประสิทธิภาพ K-nearest Neighbors

Model: K-NN	10 Fold Cross-Validation		
K	Accuracy	Precision	Recall
3	70.66%	73.06%	70.67%
4	75.83%	77.55%	75.83%
5	76.43%	77.98%	76.53%
6	76.62%	78.71%	76.69%
7	75.64%	77.59%	75.81%

4.3. ผลการหาค่าประสิทธิภาพเทคนิค Decision Tree

จากการทดลองหาค่าประสิทธิภาพโดยใช้เทคนิค Decision Tree ทำให้ได้ค่าประสิทธิภาพซึ่งประกอบไปด้วยค่าความถูกต้อง ค่าความแม่นยำ และค่าค้นคืน แสดงดังตารางที่ 5

ตารางที่ 5. แสดงเปรียบเทียบค่าประสิทธิภาพ Decision Tree

Model: DT			10 Fold Cross-Validation		
Criterion	Max Depth	Size	Accuracy	Precision	Recall
Gain ratio	10	4	64.49%	68.81%	64.72%
Gain ratio	10	2	64.49%	68.81%	64.72%
Gain ratio	20	4	70.26%	72.78%	70.30%
Infor Gain	10	4	67.26%	69.88%	67.31%
Infor Gain	10	2	67.26%	69.88%	67.31%
Infor Gain	20	4	67.45%	69.90%	67.55%
Gini	10	4	67.67%	71.22%	67.69%
Gini	10	2	67.67%	71.22%	67.69%
Gini	20	4	67.47%	70.86%	67.55%

4.4. ผลการหาค่าประสิทธิภาพเทคนิค Rule Based

จากการทดลองหาค่าประสิทธิภาพโดยใช้เทคนิค Rule Based ทำให้ได้ค่าประสิทธิภาพซึ่งประกอบไปด้วยค่าความถูกต้อง ค่าความแม่นยำ และค่าค้นคืน แสดงดังตารางที่ 6

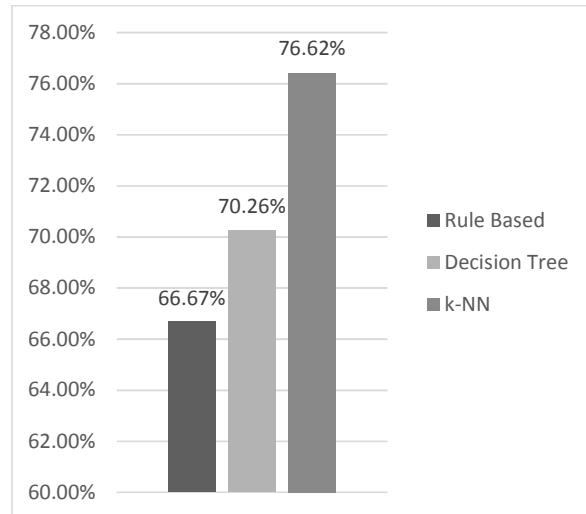
ตารางที่ 6. แสดงเปรียบเทียบค่าประสิทธิภาพ Rule Based

Model: RB				10 Fold Cross-Validation		
Criterion	Sample	Pureness	Min Prune	Accuracy	Precision	Recall
Infor Gain	0.9	0.9	0.25	66.67%	67.91%	66.81%
Infor Gain	0.8	0.8	0.3	64.67%	66.91%	64.70%
Infor Gain	1	1	0.25	61.33%	66.31%	61.60%
accuracy	0.9	0.9	0.25	58.69%	69.22%	58.89%
accuracy	0.8	0.8	0.3	61.71%	66.45%	61.90%
accuracy	1	1	0.25	57.54%	70.07%	57.87%

5. สรุปผลงานวิจัยและข้อเสนอแนะ

จากผลการดำเนินการทดสอบเปรียบเทียบประสิทธิภาพโมเดลพบว่าตัวแบบ K-nearest Neighbors มีค่าความถูกต้องมากที่สุดอยู่ที่ 76.62% รองลงมา คือ Decision Tree อยู่ที่ 70.26% และ Rule Based อยู่ที่ 66.67% ตามลำดับ แสดงดังรูปที่ 2 ซึ่งสามารถนำโมเดล K-nearest Neighbors ที่พารามิเตอร์ K = 6 ที่ได้มาใช้เป็นโมเดลในการพัฒนาระบบสนับสนุนการตัดสินใจใน

การเลือกสาขาการเรียนได้ และควรจะต้องเก็บข้อมูลต่อเนื่องของนักศึกษา ปว.ช. อย่างต่อเนื่องเพื่อจะได้ข้อมูลที่เพิ่มมากขึ้นและสามารถดูแนวโน้มได้



รูปที่ 2. แสดงการเปรียบเทียบค่าประสิทธิภาพ

6. เอกสารอ้างอิง

- [1] ขาลี ดวงแยม และคณะ. “ระบบสนับสนุนการตัดสินใจเลือกประกอบอาชีพสำหรับผู้สำเร็จการศึกษาระดับปริญญาตรีด้วยวิธีการเรียนรู้แบบเบย์”. โปรแกรมมหาวิทยาลัยการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม, 2550.
- [2] ไพฑูรย์ จันทรเรือง. “ระบบสนับสนุนการตัดสินใจเลือกสาขาการเรียนของนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคต้นไม้ตัดสินใจ”. สาขาวิชาวิทยาการคอมพิวเตอร์, ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศบัณฑิตวิทยาลัย, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2550.
- [3] N. Buniyamin, U. B. Mat, P. M. Arshad, "Educational Data Mining for Prediction and Classification of Engineering Students Achievement". International Conference on Engineering Education (ICEED), pp. 49-53, 2015
- [4] มณีรัตน์ จรุงเดชากุล. “แบบการเรียนรู้ที่มีผลต่อการสำเร็จการศึกษาของบัณฑิตคณะวิทยาศาสตร์ มหาวิทยาลัย

รามคำแหง”. งานวิจัยคณะวิทยาศาสตร์, มหาวิทยาลัย

รามคำแหง, 2554.

[5] กระทรวงศึกษาธิการ. “หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน”. พุทธศักราช 2551.

[6] Holland, L. John. “Making Vocational Choice : A Theory o Career”. New Jersey : Prentice-Hall, 1973

[7] นพวรรณ คึงชัยสกุล. “แนะนำการแปลผลแบบทดสอบความสนใจทางอาชีพตามทฤษฎีการจำแนกอาชีพตามบุคลิกภาพของ John L. Holland”. ราชิต.ร.มหาวิทยาลัยเกษตรศาสตร์ ศูนย์วิจัยและพัฒนาการศึกษา, 2551.

[8] เอกสิทธิ์ พัทธวงศ์ศักดิ์. การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไมน์นิ่ง เบื้องต้น. พิมพ์ครั้งที่ 1. กรุงเทพฯ :

บริษัท เอเชีย ดิจิตอลการพิมพ์ จำกัด, 2557.

[9] เรือนทิพย์ เข้มเพ็ชร. “การพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบวิธีต้นไม้การตัดสินใจสำหรับการเลือกสาขาวิชาเรียนของนักศึกษาระดับอุดมศึกษา”. สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยศรีปทุม, 2554.