

บทที่ 4

ผลการดำเนินงาน

4.1 ผลการวิจัยและวิจารณ์ผลการคัดคุณลักษณะ

ผลการศึกษาคูณลักษณะที่ส่งผลต่อการปัจจัยที่มีต่อการขอตำแหน่งทางวิชาการ บุคลากร กรณีศึกษาของ มหาวิทยาลัยของรัฐบาลแห่งหนึ่ง โดยใช้ เทคนิคการคำนวณหาค่า น้ำหนักของคุณลักษณะ น้ำหนัก ได้แก่ 1) Chi-Square Statistic 2) Gini Index 3) Gain Ratio 4) Information Gain และ 5) Correlation based สามารถแสดงผลค่าน้ำหนักสูงสุดที่คำนวณได้ 10 ลำดับแรกของแต่ละเทคนิค และนับจำนวนความถี่ที่ได้ 5 ลำดับแรกของแต่ละคุณลักษณะจากการใช้เทคนิคการ คัดเลือกคุณลักษณะ

ตารางที่ 4.1 แสดงคุณลักษณะ มีความสำคัญมากที่สุดของแต่ละเทคนิค

คุณลักษณะ	Correlation based	Gini Index	Gain Ratio	Information Gain	Chi- Square	ความถี่
ISI	0.709	0.082	0.193	0.193	590.453	5
TCI	0.433	0.023	0.068	0.068	196.376	5
ตำแหน่งงานปัจจุบัน	0.258	0.009	0.043	0.043	68.597	5
จำนวนผลงานอดีต	0.287	0.012	0.057	0.057	84.597	5
อายุงานปัจจุบัน	0.055	0.002	0.012	0.012	21.248	1
ระดับคุณวุฒิ	0.031	0.000	0.001	0.001	2.200	0
เขตพื้นที่	0.031	0.001	0.005	0.005	5.284	0
อายุ	0.022	0.001	0.006	0.006	12.409	0
คณะ	0.021	0.018	0.086	0.086	129.549	4
ประเภทบุคลากร	0.007	0.000	0.000	0.000	0.057	0

จากผลการวิเคราะห์ค่าความสำคัญของแอตทริบิวต์พบว่า ISI, TCI, จำนวนผลงานอดีต และตำแหน่งงานปัจจุบัน เป็นตัวแปรที่มีค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation-based) สูงสุด จึงถูกเลือกให้นำมาใช้สร้างแบบจำลอง ทั้งนี้ การเลือกใช้แอตทริบิวต์ดังกล่าวอยู่ภายใต้ขอบเขตของการวิจัยที่มุ่งวิเคราะห์ปัจจัยตามหลักเกณฑ์ของ คณะกรรมการข้าราชการพลเรือนในสถาบันอุดมศึกษา (ก.พ.อ.) ในการพิจารณาตำแหน่งทางวิชาการ โดย ISI และ TCI สะท้อนถึงคุณภาพและการยอมรับของผลงานวิจัยในระดับนานาชาติและระดับชาติซึ่งเป็นตัวชี้วัดหลักของ คณะกรรมการข้าราชการพลเรือนในสถาบันอุดมศึกษา (ก.พ.อ.) ขณะที่จำนวนผลงานอดีตแสดงถึงความต่อเนื่องและศักยภาพในการสร้างผลงานวิชาการ ส่วนตำแหน่งงานปัจจุบันเป็นเงื่อนไขเชิงโครงสร้างที่สอดคล้องกับหลักเกณฑ์การเลื่อนตำแหน่ง ดังนั้น การเลือกใช้แอตทริบิวต์ทั้งสี่ไม่เพียงแต่สอดคล้องกับผลการวิเคราะห์เชิงสถิติ แต่ยังสอดคล้องกับหลักการและกรอบเกณฑ์ที่ใช้จริงในการประเมินของ คณะกรรมการข้าราชการพลเรือนในสถาบันอุดมศึกษา (ก.พ.อ.) จึงนับว่ามีความเหมาะสมและเพิ่มความน่าเชื่อถือให้กับผลการวิจัย

4.2 ผลการประเมินโมเดลประสิทธิภาพการจำแนกข้อมูล

ตารางที่ 4.2 ผลประสิทธิภาพแต่ละโมเดลก่อนทำการปรับสมดุลข้อมูล

Model	Accuracy	Precision	Recall	F-Measure
ต้นไม้ตัดสินใจ	100.00%	100.00%	100.00%	100.00%
ป่าสุ่ม	100.00%	100.00%	100.00%	100.00%
K-NN	99.61%	100.00%	94.82%	97.14%

จากผลการทดลองตามตารางที่ 4.2 พบว่าผลประสิทธิภาพ โดยมีการนำข้อมูลจากชุดเรียนรู้มาทดสอบด้วยชุดข้อมูลทดสอบที่ให้ค่าถูกต้อง โดยใช้วิธีการ Cross-validation Test ด้วย 10-fold cross-validation เพื่อหาค่าความถูกต้อง แต่ละโมเดลก่อนปรับสมดุลข้อมูล โมเดลต้นไม้ตัดสินใจ (Decision Tree) ให้ผลการทดสอบด้วยค่า Accuracy เท่ากับ 100% ส่วนเทคนิคโมเดลป่าสุ่ม (Random Forest) และโมเดล K-NN แสดงผลลัพธ์ที่มีความแม่นยำ (Accuracy) ที่ร้อยละ 100% และ 99.61% ตามลำดับ

ตารางที่ 4.3 ผลประสิทธิภาพแต่ละโมเดลหลังทำการปรับสมดุลข้อมูล

Model	Accuracy	Precision	Recall	F-Measure
ต้นไม้ตัดสินใจ	96.01%	69.40%	100.00%	81.04%
ป่าสุ่ม	99.90%	98.75%	100.00%	99.33%
K-NN	100.00%	100.00%	100.00%	100.00%

จากผลการทดลองตามตารางที่ 4.3 พบว่าผลประสิทธิภาพ โดยมีการนำข้อมูลจากชุดเรียนรู้มาทดสอบด้วยชุดข้อมูลทดสอบที่ให้ค่าถูกต้อง โดยใช้วิธีการ Cross-validation Test ด้วย 10-fold cross-validation เพื่อหาค่าความถูกต้อง แต่ละโมเดลหลังปรับสมดุลข้อมูล โมเดลต้นไม้ตัดสินใจ (Decision Tree) ให้ผลการทดสอบด้วยค่า Accuracy เท่ากับ 96.01% ส่วนเทคนิคโมเดลป่าสุ่ม (Random Forest) และโมเดล K-NN แสดงผลลัพธ์ที่มีความแม่นยำ (Accuracy) ที่ร้อยละ 99.90% และ 100.00%ตามลำดับ

ตารางที่ 4.4 เปรียบเทียบค่าความถูกต้องก่อนและหลังการปรับสมดุลข้อมูล โดยใช้วิธีการ 10-fold cross-validation

Model	Accuracy (ก่อน)	Precision (ก่อน)	Recall (ก่อน)	FMeasure (ก่อน)	Accuracy (หลัง)	Precision (หลัง)	Recall (หลัง)	FMeasure (หลัง)
ต้นไม้ตัดสินใจ	100.00%	100.00%	100.00%	100.00%	96.01%	69.40%	100.00%	81.04%
ป่าสุ่ม	100.00%	100.00%	100.00%	100.00%	99.90%	98.75%	100.00%	99.33%
K-NN	99.61%	100.00%	94.82%	97.14%	100.00%	100.00%	100.00%	100.00%

ตามตารางที่ 4.4 ผลการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลจากการเปรียบเทียบประสิทธิภาพของแบบจำลองหลายเทคนิค พบว่าโมเดลต้นไม้ตัดสินใจ (Decision Tree) แม้จะให้ค่าตัวชี้วัดด้านความถูกต้อง (Accuracy, Recall, Precision และ F-Measure) อยู่ในระดับสูง โดยเฉพาะก่อนการปรับสมดุลข้อมูลที่มีค่าเท่ากับ 100% ทุกตัวชี้วัด แต่ผลลัพธ์ดังกล่าวสะท้อนถึงปัญหา Overfitting เนื่องจากโมเดลสามารถจดจำข้อมูลชุดฝึกได้ทั้งหมดโดยตรง ทำให้ไม่สามารถสะท้อนศักยภาพในการทำนายข้อมูลใหม่ได้อย่างแท้จริง อีกทั้งยังไม่สอดคล้องกับลักษณะของข้อมูลเชิงคุณภาพที่เกณฑ์ ก.พ.อ. ให้ความสำคัญ

ในทางตรงกันข้าม โมเดลป่าสุ่ม (Random Forest) แสดงให้เห็นถึงความสามารถในการทำนายที่ เสถียรและสมดุลมากกว่า โดยแม้ค่า Accuracy หลังการปรับสมดุลข้อมูลจะอยู่ที่

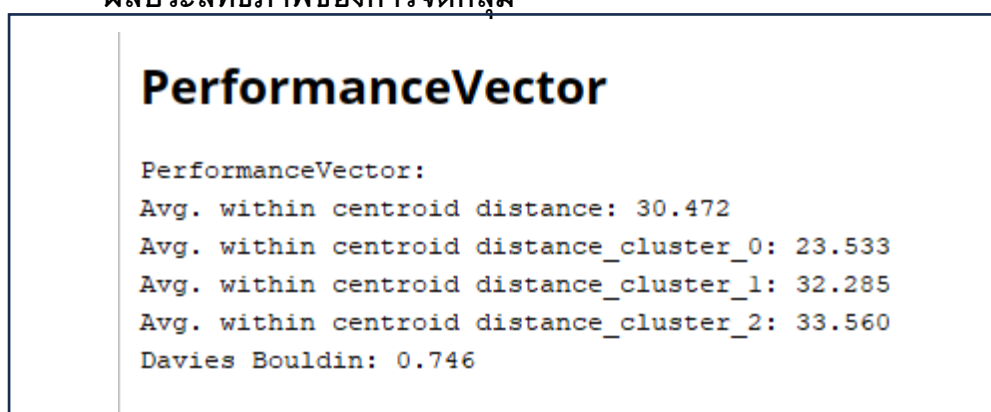
99.90% อีกทั้งวิธีการทำงานของ Random Forest ที่อาศัยการรวมผลการตัดสินใจจากต้นไม้หลายต้นยังช่วยลดความลำเอียง (Bias) และความผันผวน (Variance) ที่มักเกิดในกรณีของต้นไม้ตัดสินใจเพียงต้นเดียว

เมื่อพิจารณาพร้อมกับหลักเกณฑ์ ก.พ.อ. ซึ่งมุ่งเน้นทั้งปริมาณและคุณภาพของผลงานวิชาการ โดยเฉพาะความน่าเชื่อถือของการประเมินผล พบว่า Random Forest สามารถสะท้อนรูปแบบการเรียนรู้ได้ใกล้เคียงกับความเป็นจริงมากกว่า ทั้งในแง่ของการจำแนกบุคลากรตามความพร้อม และการหลีกเลี่ยงปัญหาการจดจำข้อมูลเฉพาะกลุ่มที่มีจำนวนมากเกินไป

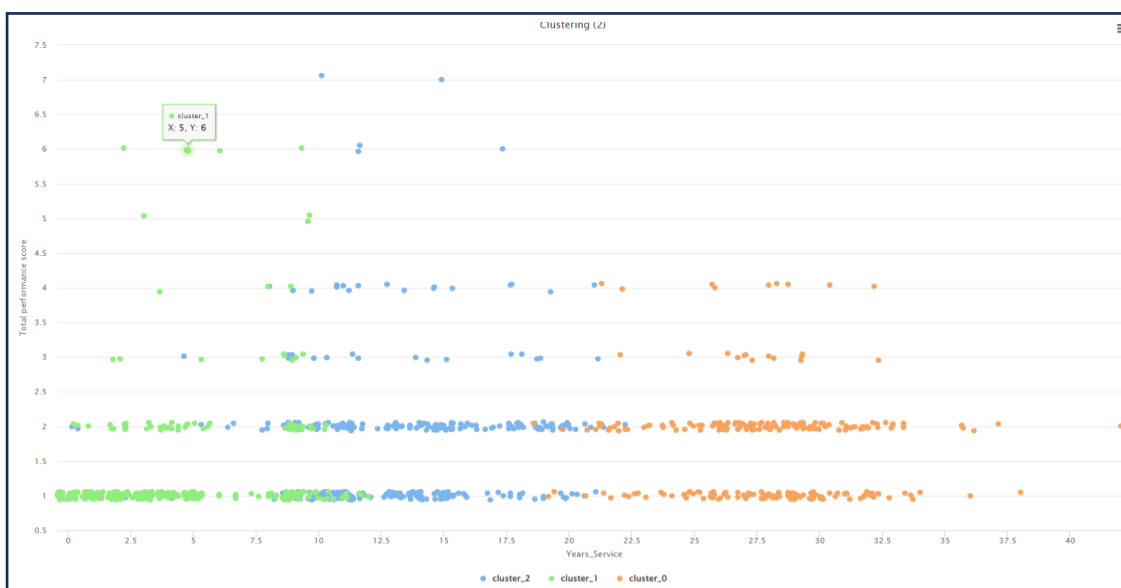
ดังนั้น ผู้จัดทำจึงได้เลือกใช้โมเดล Random Forest เป็นแบบจำลองที่เหมาะสมที่สุดสำหรับงานวิจัยครั้งนี้ เนื่องจากเป็นโมเดลที่ให้ผลการทดสอบสอดคล้องกับ เกณฑ์ ก.พ.อ. ล่าสุด และแสดงค่าตัวเลขเชิงประสิทธิภาพที่สมดุลและน่าเชื่อถือกว่าการใช้โมเดลต้นไม้ตัดสินใจเพียงต้นเดียว

4.3 สรุปผลการทำการจัดกลุ่ม

ผลประสิทธิภาพของการจัดกลุ่ม



ภาพที่ 4.1 ค่าดัชนี Davies-Bouldin Index (DBI)



ภาพที่ 4.2 ผลกลุ่ม k=3

ผลการวิเคราะห์การจัดกลุ่มข้อมูลด้วยวิธีการ K-Means เมื่อกำหนดจำนวนกลุ่มที่เหมาะสมที่สุดเป็น $K = 3$ และพล็อตข้อมูลบนกราฟสองมิติที่ใช้ตัวแปร อายุงานปัจจุบัน (Years of Service) เป็นแกน X และ คะแนนผลงานรวม (Total Performance Score) เป็นแกน Y พบว่าข้อมูลสามารถแบ่งออกเป็นสามกลุ่มที่มีลักษณะแตกต่างกันอย่างชัดเจน เป็น 3 กลุ่ม ดังนี้

กลุ่มที่ 1 Cluster 2 – กลุ่มเริ่มต้นสายอาชีพ

กลุ่มนี้กระจุกตัวอยู่บริเวณช่วงอายุงานน้อยกว่า 10-17 ปี และมีคะแนนผลงานรวมต่ำ (ประมาณ 1-2 คะแนน) แสดงให้เห็นว่าบุคลากรในกลุ่มนี้อยู่ในระยะเริ่มต้นของการสร้างผลงานวิชาการและยังอยู่ในช่วงสะสมประสบการณ์ จำเป็นต้องได้รับการพัฒนาและสนับสนุนด้านงานวิจัยอย่างต่อเนื่อง เช่น การอบรมพื้นฐานด้านการวิจัย การเขียนบทความ และการใช้ฐานข้อมูลวิชาการ เพื่อให้สามารถยกระดับศักยภาพได้ในอนาคต

กลุ่มที่ 2 Cluster 1 – กลุ่มศักยภาพเติบโตสูง

ข้อมูลของกลุ่มนี้กระจายอยู่ช่วงอายุงาน 5-15 ปี และมีคะแนนผลงานตั้งแต่ระดับกลางจนถึงสูงสุด (บางรายมีคะแนนถึง 6-7 คะแนน) ซึ่งสะท้อนถึงศักยภาพในการพัฒนาและความพร้อมในการยื่นขอตำแหน่งทางวิชาการในระยะเวลาอันใกล้ กลุ่มนี้จึงถือเป็น “กำลังหลัก” ขององค์กรทางวิชาการ ที่ควรได้รับการส่งเสริมในรูปแบบของทุนวิจัยระยะสั้น การอบรมเขียนบทความเชิงวิชาการ และการเตรียมความพร้อมเพื่อเลื่อนตำแหน่งตามเกณฑ์ของ ก.พ.อ.

กลุ่มที่ 3 Cluster 0 – กลุ่มอาวุโสประสบการณ์สูงแต่ผลงานนิ่ง

กลุ่มนี้อยู่ในช่วงอายุงานมากกว่า 20 ปี และมีคะแนนผลงานอยู่ในระดับค่อนข้างคงที่ (ประมาณ 1–3 คะแนน) สะท้อนให้เห็นถึงบุคลากรที่มีประสบการณ์และความมั่นคงในสายอาชีพ แต่มีแนวโน้มผลิตผลงานทางวิชาการลดลง การสนับสนุนควรมุ่งเน้นไปที่การกระตุ้นแรงจูงใจ เช่น การให้บทบาทพี่เลี้ยง (Mentor) การสร้างโครงการวิจัยข้ามสาขา และการส่งเสริมการตีพิมพ์ในฐานข้อมูลสากล เพื่อฟื้นฟูความมีส่วนร่วมในทางวิชาการอีกครั้ง

จึงสรุปได้ว่า ผลการจัดกลุ่มด้วย $K=3$ ทำให้สามารถจำแนกบุคลากรทางวิชาการโดยครั้งนี้ช่วยให้สามารถวางแผนเชิงนโยบายและกำหนดแนวทางพัฒนาบุคลากรได้ตรงจุดมากขึ้น โดยเฉพาะกลุ่ม **Cluster 1** ซึ่งถือเป็นกำลังหลักที่พร้อมต่อยอดสู่ตำแหน่งทางวิชาการระดับสูง ขณะที่ **Cluster 0** ควรและ **Cluster 2** ควรได้รับการส่งเสริมในลักษณะที่แตกต่างกัน เพื่อสร้างความสมดุลในการพัฒนาและความยั่งยืนของระบบอุดมศึกษาโดยรวม