

แบบรายงานผลการพัฒนาความรู้ของข้าราชการ สำนักงานพัฒนาที่ดินเขต ๒  
รอบการประเมินที่ ๒ / ๒๕๖๗ ตั้งแต่วันที่ ๑ เมษายน ๒๕๖๗ - ๓๐ กันยายน ๒๕๖๗  
ประจำปีงบประมาณ พ.ศ. ๒๕๖๗

ชื่อ-นามสกุล นางสาวดวงใจ ้วยเจริญ ตำแหน่ง นักวิชาการเกษตรชำนาญการพิเศษ  
หน่วยงาน กลุ่ม/ฝ่าย/สพด./ศูนย์ ศูนย์ศึกษาการพัฒนาเขาหินซ้อนอันเนื่องมาจากพระราชดำริ  
หัวข้อการพัฒนา การใช้โปรแกรมดิจิทัลเพื่อการวิเคราะห์ข้อมูล

วิธีการพัฒนา อบรมออนไลน์

วันที่พัฒนา ๑๕ สิงหาคม ๒๕๖๗ ถึง ๑๗ สิงหาคม ๒๕๖๗

สถานที่ ศูนย์ศึกษาการพัฒนาเขาหินซ้อนอันเนื่องมาจากพระราชดำริ

หน่วยงานที่จัดอบรม สำนักงานคณะกรรมการข้าราชการพลเรือน (สำนักงาน ก.พ.)

วัตถุประสงค์ของการเรียน

เพื่อศึกษาหาความรู้และพัฒนาทักษะเกี่ยวกับการใช้โปรแกรมดิจิทัล โปรแกรม RapidMiner เพื่อการวิเคราะห์ข้อมูล เพื่อนำมาประยุกต์ใช้งานด้านวิชาการ

### สรุปสาระสำคัญ

การใช้โปรแกรมดิจิทัลเพื่อการวิเคราะห์ข้อมูลและการสร้างโมเดลเพื่อจำแนกประเภท (Classification) โดยใช้โปรแกรม RapidMiner ซึ่งเป็นเครื่องมือที่ช่วยให้กระบวนการวิเคราะห์ข้อมูลและการสร้างโมเดลจำแนกประเภทเพื่อใช้ในการตัดสินใจสามารถทำได้อย่างรวดเร็วและมีประสิทธิภาพ ด้วยอินเทอร์เฟซที่ใช้งานง่ายและความสามารถในการลากและวางโมดูลต่าง ๆ ทำให้ผู้ใช้สามารถสร้างกระบวนการวิเคราะห์ที่ซับซ้อนได้โดยไม่ต้องเขียนโค้ดเอง การสร้างโมเดลเพื่อจำแนกประเภท (Classification Model) โดยใช้ RapidMiner เป็นเทคนิคในการวิเคราะห์ข้อมูลที่ใช้ในการจำแนกกลุ่มข้อมูลออกเป็นประเภทต่าง ๆ โดยใช้โมเดลการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งมีขั้นตอนการทำงาน ดังต่อไปนี้

#### ๑. การนำเข้าข้อมูล (Data Import)

โปรแกรม RapidMiner รองรับการนำเข้าข้อมูลจากหลายรูปแบบ เช่น ไฟล์ CSV, Excel, ฐานข้อมูล (SQL), ข้อมูลจากเว็บเซอร์วิส และอื่น ๆ โดยใช้โมดูล "Read CSV", "Read Excel", หรือ "Retrieve" เพื่อดึงข้อมูลเข้ามาในกระบวนการ (Process)

#### ๒. การสำรวจข้อมูล (Data Exploration)

ตรวจสอบข้อมูลเบื้องต้นเพื่อทำความเข้าใจโครงสร้างของข้อมูล เช่น การดูค่าต่าง ๆ ที่ปรากฏในแต่ละคอลัมน์, การเช็คค่าที่ขาดหาย (missing values), การหาค่าที่เป็น outliers เป็นต้น ใช้โมดูล "Statistics" หรือ "Data Overview" เพื่อดูสถิติพื้นฐานของข้อมูล เช่น ค่าเฉลี่ย, ค่ามัธยฐาน, ค่ามากที่สุดและน้อยสุด เป็นต้น

#### ๓. การจัดการค่าที่ขาดหาย (Handling Missing Values)

ค่าที่ขาดหาย (missing values) ในข้อมูลสามารถส่งผลกระทบต่อกระบวนการวิเคราะห์ได้ จึงต้องจัดการก่อน RapidMiner มีโมดูล "Replace Missing Values" ซึ่งสามารถใช้แทนค่าที่ขาดหายด้วยค่าที่เหมาะสม เช่น ค่าเฉลี่ย (mean), ค่ามัธยฐาน (median), ค่าที่พบบ่อย (mode) หรือใช้วิธีการลบแถวที่มีค่าขาดหายออกไป

#### ๔. การกรองข้อมูล (Data Filtering)

บางครั้งข้อมูลที่ไม่เกี่ยวข้องหรือมีปัญหาควรถูกรองออกไปก่อนการวิเคราะห์ โดยใช้โมดูล "Filter Examples" หรือ "Filter Examples Range" เพื่อกรองข้อมูลตามเงื่อนไขที่ต้องการ เช่น กรองแถวที่มีค่าเฉพาะ, ลบข้อมูลที่อยู่นอกช่วงค่าที่กำหนด เป็นต้น

#### ๕. การแปลงข้อมูล (Data Transformation)

แปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับการวิเคราะห์ เช่น การเปลี่ยนค่าตัวเลขให้เป็นกลุ่ม (binning), การสร้างฟีเจอร์ใหม่จากฟีเจอร์เดิม, การรวมคอลัมน์ หรือแยกคอลัมน์ ใช้โมดูล "Normalize" เพื่อปรับสเกลของข้อมูลให้อยู่ในช่วงที่กำหนด, โมดูล "Generate Attributes" เพื่อสร้างฟีเจอร์ใหม่, หรือ "Pivot" เพื่อจัดการกับข้อมูลที่อยู่ในรูปแบบ pivot

#### ๖. การจัดการกับข้อมูลที่ซ้ำซ้อน (Handling Duplicates)

ข้อมูลที่ซ้ำซ้อนอาจทำให้การวิเคราะห์เกิดความผิดพลาดได้ จึงต้องตรวจสอบและลบออกไป ใช้โมดูล "Remove Duplicates" เพื่อลบแถวที่มีค่าซ้ำกันในชุดข้อมูล

#### ๗. การสร้างตัวแปรที่เหมาะสม (Feature Engineering)

สร้างตัวแปรใหม่ที่ช่วยให้การวิเคราะห์มีประสิทธิภาพขึ้น เช่น การคำนวณค่าเฉลี่ยของหลายคอลัมน์, การสร้างตัวแปรที่บ่งบอกถึงกลุ่ม หรือการรวมข้อมูลจากหลายแหล่งเข้าด้วยกัน ใช้โมดูล "Generate Attributes" หรือ "Polynomial Expansion" เพื่อสร้างฟีเจอร์ใหม่

#### ๘. การแบ่งชุดข้อมูล (Data Splitting)

ในกรณีที่ต้องการเตรียมข้อมูลสำหรับการสร้างโมเดล ควรแบ่งข้อมูลออกเป็นชุดฝึกอบรม (Training Data) และชุดทดสอบ (Test Data) ใช้โมดูล "Split Data" หรือ "Cross Validation" เพื่อแบ่งข้อมูลตามสัดส่วนที่ต้องการ

#### ๙. การตรวจสอบข้อมูลสุดท้าย (Final Data Check)

ตรวจสอบข้อมูลที่เตรียมไว้ก่อนนำไปใช้ในการวิเคราะห์หรือสร้างโมเดล เพื่อให้แน่ใจว่าข้อมูลมีความสมบูรณ์และพร้อมใช้งาน ใช้โมดูล "Data Overview" หรือ "Statistics" เพื่อตรวจสอบข้อมูลครั้งสุดท้าย

#### ๑๐. การสร้างต้นไม้ตัดสินใจ (Decision Tree)

เป็นวิธีการที่มีประสิทธิภาพในการจำแนกประเภทข้อมูลและการสร้างโมเดลการเรียนรู้ของเครื่อง (Machine Learning Model) ขั้นตอนในการสร้างต้นไม้ตัดสินใจใน RapidMiner มีขั้นตอนดังนี้ คือ ลากและวางโมดูล "Decision Tree" จากส่วนของ Operators ใน RapidMiner ไปยังพื้นที่ทำงาน (Process) ตั้งค่าพารามิเตอร์ ของโมดูล "Decision Tree" ตามความต้องการ เช่น ความลึกสูงสุดของต้นไม้ (max depth), การตั้งค่า criterion สำหรับการแบ่ง (เช่น Gini Index หรือ Entropy) และการตั้งค่าต่าง ๆ ที่เกี่ยวข้อง

#### ๑๑. การฝึกอบรมโมเดล (Model Training)

เชื่อมต่อโมดูล "Decision Tree" กับชุดข้อมูลฝึกอบรม (Training Data) และใช้โมดูล "Apply Model" เพื่อฝึกอบรมโมเดลด้วยข้อมูลที่มีอยู่



**๑๒. การประเมินผลโมเดล (Model Evaluation)**

ใช้โมดูล "Apply Model" และ "Performance" เพื่อตรวจสอบประสิทธิภาพของโมเดลต้นไม้ตัดสินใจ ด้วยชุดข้อมูลทดสอบ (Test Data) ดูผลลัพธ์ เช่น Accuracy, Precision, Recall, F๑ Score และ Confusion Matrix เพื่อประเมินความแม่นยำและประสิทธิภาพของโมเดล

**๑๓. การปรับแต่งโมเดล (Model Tuning)**

ปรับแต่งพารามิเตอร์ของโมเดลต้นไม้ตัดสินใจเพื่อเพิ่มประสิทธิภาพของโมเดล โดยอาจทดลองกับค่า parameters ที่แตกต่างกันหรือใช้เทคนิคการเลือกคุณสมบัติที่ดีที่สุด ใช้ "Parameter Optimization" หรือ "Cross Validation" เพื่อปรับแต่งและประเมินโมเดล

**๑๔. การนำเสนอผลลัพธ์ (Results Presentation)**

ใช้โมดูล "Decision Tree Visualization" เพื่อแสดงภาพของต้นไม้ตัดสินใจในรูปแบบกราฟฟิก ซึ่งช่วยให้เห็นโครงสร้างของต้นไม้และการตัดสินใจที่เกิดขึ้นได้ง่ายขึ้น สร้างรายงานหรือแดชบอร์ดที่แสดงผลลัพธ์ของการจำแนกประเภทเพื่อการนำเสนอและนำไปใช้ในการตัดสินใจได้อย่างแม่นยำ

**ประโยชน์ที่ได้รับจากการพัฒนาความรู้**

- ๑. ทำให้สามารถวิเคราะห์ข้อมูลที่มีเพื่อการตัดสินใจอย่างมีประสิทธิภาพเพิ่มขึ้น
- ๒. พัฒนาทักษะในการจัดการและวิเคราะห์ข้อมูล ที่สามารถนำมาประยุกต์ใช้กับการทำงานวิจัยได้

การนำองค์ความรู้ไปประยุกต์ใช้ สามารถนำองค์ความรู้ที่ได้รับไปประยุกต์ใช้ในการวิเคราะห์ข้อมูลในการศึกษา ทดลอง วิจัยด้านการเกษตร เช่น การวิเคราะห์ข้อมูลเพื่อการทำนายผลผลิตการเกษตร เป็นต้น

**แนวทางการถ่ายทอดองค์ความรู้สู่เกษตรกร**

จัดกิจกรรมการเรียนรู้ หรือฝึกอบรมให้กับเกษตรกรรุ่นใหม่ สามารถใช้โปรแกรม RapidMiner เป็นเครื่องมือในการวิเคราะห์ข้อมูลและช่วยตัดสินใจในการวางแผนการผลิต

(ลงนาม) ..... 

(นางสาวดวงใจ วยเจริญ)  
นักวิชาการเกษตรชำนาญการพิเศษ

(ลงนาม) ..... 

(นายอนุรักษ์ บัวคลี่คล้าย)  
รักษาการในตำแหน่ง ผู้อำนวยการศูนย์ศึกษาการพัฒนาเขาหินซ้อน  
อันเนื่องมาจากพระราชดำริ