

แบบรายงานผลการพัฒนาความรู้ของข้าราชการ สำนักงานพัฒนาที่ดินเขต ๒  
รอบการประเมินที่ ๑ / ๒๕๖๘ ตั้งแต่วันที่ ๑ ตุลาคม ๒๕๖๗ - ๓๑ มีนาคม ๒๕๖๘  
ประจำปีงบประมาณ พ.ศ. ๒๕๖๘

ชื่อ-นามสกุล นางสาวสมิทธา มงคล ตำแหน่ง นักวิชาการเกษตรชำนาญการพิเศษ  
หน่วยงาน กลุ่มวิชาการเพื่อการพัฒนาที่ดิน สำนักงานพัฒนาที่ดินเขต ๒  
หัวข้อการพัฒนา การใช้โปรแกรมดิจิทัลเพื่อการวิเคราะห์ข้อมูล  
วิธีการพัฒนา เข้าร่วมการพัฒนาด้วยระบบสื่ออิเล็กทรอนิกส์ (e-Learning)  
วันที่พัฒนา ๑๔ มกราคม ๒๕๖๘  
สถานที่ กลุ่มวิชาการเพื่อการพัฒนาที่ดิน สำนักงานพัฒนาที่ดินเขต ๒  
หน่วยงานที่จัดอบรม สำนักงานคณะกรรมการข้าราชการพลเรือน  
วัตถุประสงค์ของการเรียน

เพื่อให้มีความรู้และความเข้าใจเกี่ยวกับการใช้โปรแกรมดิจิทัลเพื่อการวิเคราะห์ข้อมูลได้อย่างถูกต้อง  
สรุปสาระสำคัญ

เป็นเนื้อหาที่ทำให้ทราบถึงภาพรวมของการใช้โปรแกรมดิจิทัลเพื่อการวิเคราะห์ข้อมูล การเตรียม  
ข้อมูล และการสร้างโมเดลเพื่อจำแนกประเภทโดยใช้โปรแกรม RapidMiner ตามรายละเอียดตามเอกสาร  
ที่แนบมาท้ายนี้

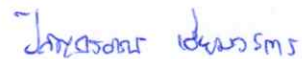
ประโยชน์ที่ได้รับจากการพัฒนาความรู้

สามารถนำความรู้เกี่ยวกับการใช้โปรแกรมดิจิทัลเพื่อการวิเคราะห์ข้อมูล มาปรับใช้ในการปฏิบัติงาน  
ทางด้านวิชาการได้อย่างมีประสิทธิภาพ สามารถใช้โปรแกรม RapidMiner มาประยุกต์ใช้ในการตัดสินใจ  
การศึกษา ทดลอง การวิเคราะห์ข้อมูลได้อย่างถูกต้อง และช่วยตัดสินใจในการวางแผนการผลิต



(นางสาวสมิทธา มงคล)

นักวิชาการเกษตรชำนาญการพิเศษ



(นางสาววลัญชรชน เอี่ยมวรกร)

ผู้อำนวยการกลุ่มวิชาการเพื่อการพัฒนาที่ดิน



## การใช้โปรแกรมดิจิทัลเพื่อการวิเคราะห์ข้อมูล

การใช้โปรแกรมดิจิทัลเพื่อการวิเคราะห์ข้อมูล และการสร้างโมเดลเพื่อจำแนกประเภทโดยใช้โปรแกรม RapidMiner เป็นเครื่องมือที่ช่วยในกระบวนการวิเคราะห์ข้อมูล และสร้างโมเดลจำแนกประเภทเพื่อใช้ในการตัดสินใจสามารถทำได้อย่างรวดเร็ว มีประสิทธิภาพ ใช้งานง่ายด้วยอินเทอร์เฟซแบบลากและวางโมดูลต่างๆ ทำให้ผู้ใช้สามารถสร้างกระบวนการวิเคราะห์ที่ซับซ้อนได้โดยไม่ต้องเขียนโค้ดเอง การสร้างโมเดลเพื่อจำแนกประเภท (Classification Model) โดยใช้ RapidMiner เป็นเทคนิคในการวิเคราะห์ข้อมูลที่ใช้ในการจำแนกกลุ่มข้อมูลออกเป็นประเภทต่างๆ โดยใช้โมเดลการเรียนรู้ของเครื่อง (Machine Learning)

**กระบวนการทำงานของ โปรแกรม RapidMiner** มีดังต่อไปนี้

### ๑. การนำเข้าข้อมูล (Data Import)

โปรแกรม RapidMiner รองรับการนำเข้าข้อมูลจากหลายรูปแบบ เช่น ไฟล์ CSV, Excel, ฐานข้อมูล SQL, ข้อมูลจากเว็บเซอร์วิส และอื่นๆ โดยใช้โมดูล Read CSV, Read Excel, Retrieve เพื่อดึงข้อมูลเข้ามาในกระบวนการ (Process)

### ๒. การสำรวจข้อมูล (Data Exploration)

ตรวจสอบข้อมูลเบื้องต้นเพื่อทำความเข้าใจโครงสร้างของข้อมูล เช่น การดูค่าต่างๆ ที่ปรากฏในแต่ละคอลัมน์, การตรวจสอบค่าที่ขาดหาย (missing values), outliers, ค่าสถิติต่างๆ เป็นต้น โดยใช้โมดูล Statistics หรือ Data Overview เพื่อดูสถิติพื้นฐานของข้อมูล เช่น ค่าเฉลี่ย (mean), ค่ามัธยฐาน (median) ค่ามากที่สุดและน้อยสุด เป็นต้น

### ๓. การจัดการค่าที่ขาดหาย (Handling Missing Values)

ค่าที่ขาดหาย (Missing values) ในข้อมูลสามารถส่งผลกระทบต่อกระบวนการวิเคราะห์ได้ จึงต้องจัดการก่อน โดยใช้โมดูล Replace Missing Values ซึ่งสามารถใช้แทนค่าที่ขาดหายด้วยค่าที่เหมาะสม เช่น ค่าเฉลี่ย (mean), ค่ามัธยฐาน (median), ค่าที่พบบ่อย (Mode) หรือใช้วิธีการลบข้อมูลที่มีค่าว่าง

### ๔. การกรองข้อมูล (Data Filtering)

บางครั้งข้อมูลที่ไม่เกี่ยวข้องหรือมีปัญหาควรถูกกำจัดออกไปก่อนการวิเคราะห์ โดยใช้โมดูล Filter Examples เพื่อกรองข้อมูลตามเงื่อนไขที่ต้องการ เช่น ค่าที่ผิดปกติ

### ๕. การแปลงข้อมูล (Data Transformation)

แปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับการวิเคราะห์ เช่น การแปลงค่าตัวเลขเป็นกลุ่ม (binning) การสร้างฟีเจอร์ใหม่ การปรับสเกลข้อมูล โดยใช้โมดูล Normalize เพื่อปรับสเกลของข้อมูลให้อยู่ในช่วงที่กำหนด ใช้โมดูล Generate Attributes เพื่อสร้างฟีเจอร์ใหม่ หรือ โมดูล Pivot เพื่อจัดการกับข้อมูลที่อยู่ในรูปแบบ pivot

### ๖. การจัดการกับข้อมูลที่ซ้ำซ้อน (Handling Duplicates)

ข้อมูลที่ซ้ำซ้อนอาจทำให้การวิเคราะห์เกิดความผิดพลาดได้ จึงต้องตรวจสอบและลบออกไป โดยใช้โมดูล Remove Duplicates เพื่อลบแถวที่มีค่าซ้ำกันในชุดข้อมูล

### ๗. การสร้างตัวแปรที่เหมาะสม (Feature Engineering)

การสร้างตัวแปรใหม่ที่ช่วยให้การวิเคราะห์มีประสิทธิภาพขึ้น เช่น การคำนวณค่าเฉลี่ยของหลายคอลัมน์ การสร้างตัวแปรที่บ่งบอกถึงกลุ่ม หรือการรวมข้อมูลจากหลายแหล่งเข้าด้วยกันโดยใช้โมดูล Generate Attributes หรือ Polynomial Expansion เพื่อสร้างฟีเจอร์ใหม่



## ๘. การแบ่งชุดข้อมูล (Data Splitting)

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

## ๙. การตรวจสอบข้อมูลสุดท้าย (Final Data Check)

ตรวจสอบข้อมูลที่เตรียมไว้ก่อนนำไปใช้ในการวิเคราะห์หรือสร้างโมเดล เพื่อให้แน่ใจว่าข้อมูลมีความสมบูรณ์และพร้อมใช้งานใช้โมดูล Data Overview หรือ Statistics เพื่อตรวจสอบข้อมูลครั้งสุดท้าย

## การสร้างโมเดลจำแนกประเภท (Classification Model)



### ๑. การสร้างต้นไม้ตัดสินใจ (Decision Tree)

เป็นวิธีการที่มีประสิทธิภาพในการจำแนกประเภทข้อมูลและการสร้างโมเดลการเรียนรู้ของเครื่อง (Machine learning Model) การสร้างต้นไม้ตัดสินใจใน RapidMiner มีขั้นตอนดังนี้ คือ ลากและวางโมเดล Decision Tree จากส่วนของ Operators ใน RapidMiner ไปยังพื้นที่ทำงาน (Process) ตั้งค่าพารามิเตอร์ของโมดูล Decision Tree ตามความต้องการ เช่น ตั้งค่าความลึกสูงสุดของต้นไม้ (Max Depth) การตั้งค่า criterion สำหรับการแบ่ง เช่น Gini Index หรือ Entropy และการตั้งค่าต่างๆ ที่เกี่ยวข้อง

### ๒. การฝึกอบรมโมเดล (Model Training)

การเชื่อมต่อโมดูล Decision Tree กับชุดข้อมูลฝึกอบรม (Training Data) และการใช้โมดูล Apply Model เพื่อฝึกอบรมโมเดลด้วยข้อมูลที่มีอยู่

### ๓. การประเมินผลโมเดล (Model Evaluation)

ใช้โมดูล Performance เพื่อตรวจสอบประสิทธิภาพของโมเดลต้นไม้ตัดสินใจด้วยชุดข้อมูลทดสอบ (Test Data) ดูผลลัพธ์ เช่น Accuracy Precision Recall F๑Score และ Confusion Matrix เพื่อประเมินความแม่นยำ และประสิทธิภาพของโมเดล

### ๔. การปรับแต่งโมเดล (Model Tuning)

ทดลองปรับค่าพารามิเตอร์ของโมเดลต้นไม้ตัดสินใจเพื่อเพิ่มประสิทธิภาพ อาจทดลองกับค่า Parameters ที่แตกต่างกัน หรือใช้เทคนิคการเลือกคุณสมบัติที่ดีที่สุด โดยใช้โมดูล Parameters Optimization หรือ Cross Validation เพื่อปรับแต่งและประเมินโมเดล

### ๕. การนำเสนอผลลัพธ์ (Results Presentation)

แสดงผลลัพธ์โดยใช้โมดูล Decision Tree Visualization เพื่อแสดงภาพของต้นไม้ตัดสินใจในรูปแบบกราฟิก ซึ่งช่วยให้เห็นโครงสร้างของต้นไม้และการตัดสินใจที่เกิดขึ้นได้ง่ายขึ้น สร้างรายงาน ตารางสรุป หรือแดชบอร์ดที่แสดงผลลัพธ์ของการจำแนกประเภทเพื่อการนำเสนอ และนำไปใช้ในการตัดสินใจได้อย่างแม่นยำ

ดังนั้น โปรแกรม RapidMiner ช่วยให้การวิเคราะห์ข้อมูลและการสร้างโมเดล Machine Learning ทำได้ง่ายขึ้น ด้วยการลากและวางโมดูลแบบไม่ต้องเขียนโค้ด รองรับการเตรียมข้อมูล สร้างโมเดล ทดสอบ และปรับแต่ง เพื่อให้ได้ผลลัพธ์ที่แม่นยำที่สุด เหมาะสำหรับการวิเคราะห์เชิงธุรกิจ การคาดการณ์ และการตัดสินใจเชิงกลยุทธ์



