

สรุปบทเรียนการพัฒนาความรู้
หลักสูตร การใช้โปรแกรมดิจิทัลเพื่อการวิเคราะห์ข้อมูล
หน่วยงาน สำนักงานคณะกรรมการข้าราชการพลเรือน (สำนักงาน ก.พ.)

ชื่อ-นามสกุล: นางสาวอภินันท์ พุกพงษ์ ตำแหน่ง: นักวิชาการเกษตรชำนาญการพิเศษ
หน่วยงาน: กลุ่มวิจัยและพัฒนาการปรับปรุงบำรุงดิน กองวิจัยและพัฒนาการจัดการที่ดิน
วันที่อบรม: ๑๙-๒๐ กุมภาพันธ์ ๒๕๖๘

๑.วัตถุประสงค์ของหลักสูตร

เพื่อศึกษาหาความรู้และพัฒนาทักษะเกี่ยวกับการใช้โปรแกรมดิจิทัล โปรแกรม RapidMiner เพื่อการวิเคราะห์ข้อมูล เพื่อนำมาประยุกต์ใช้งานด้านวิชาการ

๒.สรุปบทเรียน

โปรแกรม RapidMiner เป็นเครื่องมือที่ช่วยให้กระบวนการวิเคราะห์ข้อมูล Data Science สามารถทำงานได้โดยไม่ต้องมีพื้นฐานการเขียนโปรแกรม มีขั้นตอนที่ง่ายและมีประสิทธิภาพ ตั้งแต่การการเก็บข้อมูล การทำความสะอาด การเตรียมข้อมูล การวิเคราะห์ ไปจนถึงการสร้างโมเดล ประเมินผล มีเทคนิคการสร้างโมเดลจำแนกประเภทข้อมูล โดยใช้การเรียนรู้ของเครื่อง (Machine Learning) รวมทั้งการงานจริงแบบเรียลไทม์ เพื่อค้นหา ข้อมูลเชิงลึก (Insights) ที่มีประโยชน์สำหรับการตัดสินใจหรือแก้ปัญหาเฉพาะด้าน

๑.โปรแกรม RapidMiner มีโครงสร้างการใช้งานประกอบด้วย

๑.๑ RapidMiner Studio เป็นเครื่องมือวิเคราะห์ข้อมูลบนเดสก์ท็อปที่ช่วยให้ผู้ใช้สามารถสร้างแบบจำลองการวิเคราะห์ข้อมูล, ประมวลผลข้อมูล, และนำเสนอข้อมูลในรูปแบบต่างๆ RapidMiner Studio มีคุณสมบัติหลากหลาย เช่น การสร้างแบบจำลองการวิเคราะห์ข้อมูล, การประมวลผลข้อมูล, การนำเสนอข้อมูล, และการเชื่อมต่อกับฐานข้อมูลต่างๆ

๑.๒ RapidMiner Server เป็นเครื่องมือจัดการและเผยแพร่แบบจำลองการวิเคราะห์ข้อมูลบนเซิร์ฟเวอร์ RapidMiner Server ช่วยให้ผู้ใช้สามารถเผยแพร่แบบจำลองการวิเคราะห์ข้อมูล, จัดการการเข้าถึงแบบจำลอง, และติดตามผลการทำงานของแบบจำลอง RapidMiner Server ยังช่วยให้ผู้ใช้สามารถเชื่อมต่อกับฐานข้อมูลต่างๆ และเผยแพร่แบบจำลองการวิเคราะห์ข้อมูลไปยังแอปพลิเคชันต่างๆ

๑.๓ RapidMiner AI Hub เป็นแพลตฟอร์มการวิเคราะห์ข้อมูลและปัญญาประดิษฐ์ที่ช่วยให้ผู้ใช้สามารถสร้าง, จัดการ, และเผยแพร่แบบจำลองการวิเคราะห์ข้อมูลและปัญญาประดิษฐ์ RapidMiner AI Hub มีคุณสมบัติหลากหลาย เช่น การสร้างแบบจำลองการวิเคราะห์ข้อมูลและปัญญาประดิษฐ์, การจัดการแบบจำลอง, การเผยแพร่แบบจำลอง, และการเชื่อมต่อกับฐานข้อมูลต่างๆ

๑.๔ RapidMiner Go เป็นแอปพลิเคชันบนเว็บที่ช่วยให้ผู้ใช้สามารถสร้าง, จัดการ, และเผยแพร่แบบจำลองการวิเคราะห์ข้อมูลและปัญญาประดิษฐ์ RapidMiner Go มีคุณสมบัติหลากหลาย เช่น การสร้างแบบจำลองการวิเคราะห์ข้อมูลและปัญญาประดิษฐ์, การจัดการแบบจำลอง, การเผยแพร่แบบจำลอง, และการเชื่อมต่อกับฐานข้อมูลต่างๆ RapidMiner Go ยังช่วยให้ผู้ใช้สามารถทำงานร่วมกับผู้อื่นและติดตามผลการทำงานของแบบจำลองได้

๒. รูปแบบการใช้งานของโปรแกรม RapidMiner มีส่วนประกอบ ดังนี้

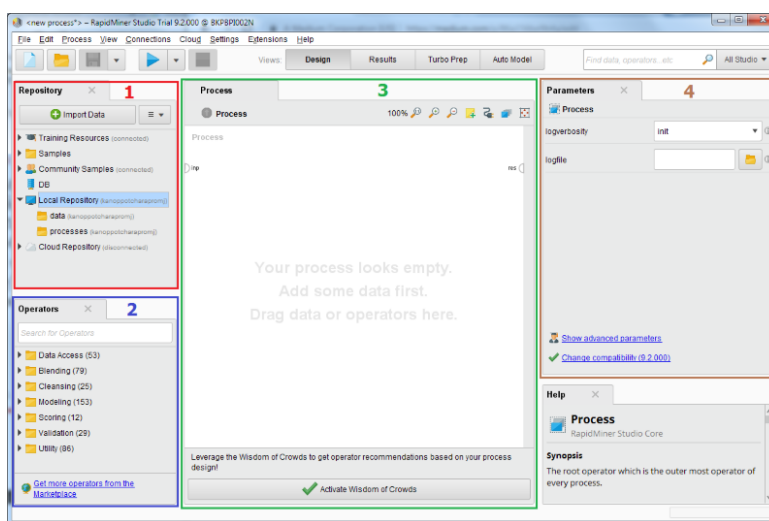
๒.๑ Repository : ส่วนนี้จะใช้ในการจัดการไฟล์ต่างๆ ของRapidMiner โดยจะเก็บไฟล์ข้อมูล หรือ Process ต่างๆไว้ใน Folder เพื่อความสะดวกในการเรียกใช้งานครั้งถัดไป

๒.๒ Operators : ส่วนนี้จะเก็บโอเพอเรเตอร์ (Operators) ในการใช้งานต่างๆ ไว้เป็นกลุ่ม ซึ่งสามารถค้นหา Operators ที่ต้องการได้ในช่อง Search เพื่อความสะดวกในการเรียกใช้งาน

๒.๓ Process : ส่วนนี้เป็นส่วนที่สำคัญ ของ RapidMiner เพราะเป็นการนำเอา Operators ต่างๆ มาเชื่อมกันให้ เป็น Process ขึ้นมาใช้งานในการทางานเกี่ยวกับการเรียนรู้ของเครื่อง (Machine Learning)

๒.๔ Parameters : ส่วนนี้จะเป็นส่วนที่แสดงพารามิเตอร์(parameter) ที่เกี่ยวข้องกับแต่ละ Operator เพื่อให้ผู้ใช้งานปรับแต่งตามที่ต้องการ

๒.๕ Help: จะเป็นคำอธิบายของเครื่องมือที่มาวางไว้ในProcess

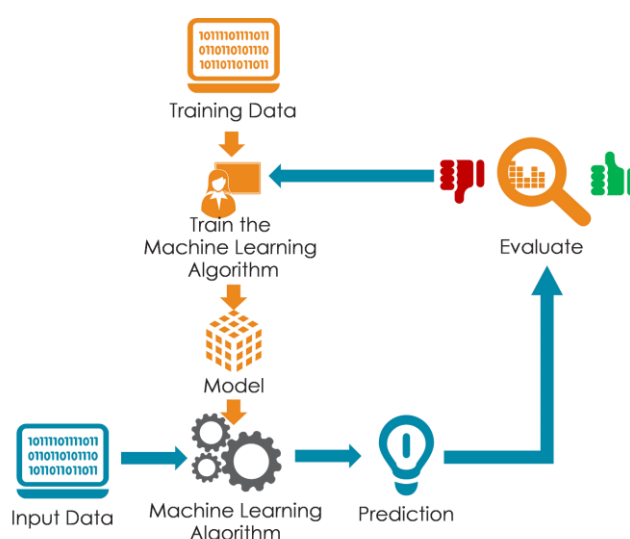


รูปที่ ๑ ส่วนประกอบของโปรแกรม RapidMiner

๓.วิธีการทำงานของโปรแกรม RapidMiner

โปรแกรม RapidMiner ทำงานในรูปแบบของ Workflow-Based Data Science ซึ่งหมายถึงการสร้างกระบวนการวิเคราะห์ข้อมูลผ่าน การลากและวางโมดูล (Operators) ที่เชื่อมต่อกันเป็นเครือข่าย (Process Flow) โดยทั่วไป มีขั้นตอนการทำงานหลัก ๕ ขั้นตอน คือ

๑. การนำเข้าข้อมูล (Data Import)
๒. การทำความสะอาดและเตรียมข้อมูล (Data Preprocessing)
๓. การสร้างและฝึกโมเดล (Modeling & Machine Learning)
๔. การประเมินผลโมเดล (Model Evaluation)
๕. การนำไปใช้งานจริง (Deployment)



รูปที่ ๒ ขั้นตอนการทำงานของโปรแกรม RapidMiner

ขั้นตอนการทำงานของโปรแกรม RapidMiner มีรายละเอียดดังนี้

๓.๑ การนำเข้าข้อมูล (Data Import)

RapidMiner รองรับการนำเข้าข้อมูลจากหลายแหล่ง เช่น: ไฟล์: CSV, Excel, JSON, XML, ฐานข้อมูล: MySQL, PostgreSQL, SQL Server, Oracle, Big Data: Apache Hadoop, Spark, API/Web: REST API, Web Scraping, Cloud Storage: Google Drive, AWS S๓

วิธีการนำเข้าข้อมูล

- ๑) เปิด RapidMiner Studio
- ๒) เลือก Operators : Retrieve เพื่อนำเข้าข้อมูลจากภายใน RapidMiner
- ๓) เลือก Read CSV / Read Excel เพื่อนำเข้าไฟล์
- ๔) ใช้ Database Connector หากต้องการเชื่อมต่อฐานข้อมูล
- ๕) ใช้ Web Mining หากต้องการดึงข้อมูลจากเว็บไซต์

๓.๒ การทำความสะอาดและเตรียมข้อมูล (Data Preprocessing)

Data คือข้อมูลที่อยู่ในตาราง ประกอบไปด้วยแถวและคอลัมน์ โดยข้อมูลในแถวเรียกว่า ตัวอย่าง (example) ข้อมูลในคอลัมน์ เรียกว่า แอททริบิวต์ (Attribute) มีหน้าที่ (role) ๓ แบบ คือ

- ๑) ไอดี (ID) เป็นแอตทริบิวต์ที่แสดงหมายเลขของข้อมูลหรือ primary key ในฐานข้อมูล
- ๒) แอททริบิวต์ทั่วไป (attribute) เป็นแอตทริบิวต์ปกติที่จะใช้ในการสร้างโมเดลหรือเรียกว่า เป็นฟีเจอร์ (feature) หรือตัวแปรต้น (independent variable)
- ๓) ลาเบล (label) เป็นแอตทริบิวต์ชนิดพิเศษที่มักจะใช้แสดงคำตอบของสิ่งที่เราต้องการจะสร้างโมเดลทำนาย หรือเรียกว่าคลาส (class) หรือตัวแปรตาม (dependent variable)

ประเภทของข้อมูลแบ่งเป็น

- ๑) Nominal ข้อมูลประเภท category (ข้อมูลที่ไม่ใช่ตัวเลข) มีค่ามากกว่า ๒ ค่าขึ้นไป
- ๒) Binomial ข้อมูลประเภท category (ข้อมูลที่ไม่ใช่ตัวเลข) มีค่าเพียง ๒ ค่าเท่านั้น
- ๓) Numeric ข้อมูลประเภทตัวเลข
- ๔) Text ข้อมูลประเภทข้อความ

ก่อนนำข้อมูลไปใช้วิเคราะห์ ต้องทำการจัดการข้อมูลให้อยู่ในรูปแบบที่เหมาะสม โมดูลที่ช่วยจัดการข้อมูล เช่น:

- Handling Missing Values : กรอกราคาเฉลี่ย, ลบแถวที่มีค่า missing
- Normalization & Standardization : ปรับค่าข้อมูลให้อยู่ในช่วงที่เหมาะสม
- Feature Selection : เลือกเฉพาะฟีเจอร์ที่สำคัญ
- Data Transformation : ใช้ Operator Numerical to Binomial แปลงประเภทข้อมูล เช่น ตัวเลขเป็นหมวดหมู่
- ลบข้อมูลซ้ำ: ใช้ Operator Remove Duplicates เพื่อลบข้อมูลที่ซ้ำกันออก
- Replace Missing Values : กำหนดค่าแทนที่ข้อมูลที่หายไปด้วยค่าเฉลี่ย (Mean) หรือค่ากลาง (Median)
- Filter Examples : คัดกรองเฉพาะข้อมูลที่ต้องการ
- Set Role : กำหนดว่าแต่ละฟีเจอร์เป็น Attribute หรือ Label

การวิเคราะห์ข้อมูลเบื้องต้น (Exploratory Data Analysis: EDA)

ขั้นตอนนี้เป็น การสำรวจข้อมูลเพื่อทำความเข้าใจข้อมูลเบื้องต้น เช่น การกระจายตัวของข้อมูล, ความสัมพันธ์ระหว่างตัวแปร เช่น

- **สร้างกราฟ:** ใช้ Operator Plot View เพื่อสร้างกราฟ Histogram, Scatter Plot หรือ Box Plot เพื่อดูการกระจายตัวของข้อมูล
- **สถิติเบื้องต้น:** ใช้ Operator Statistics เพื่อดูค่าเฉลี่ย, ส่วนเบี่ยงเบนมาตรฐาน, ค่าสูงสุด และต่ำสุดของแต่ละตัวแปร

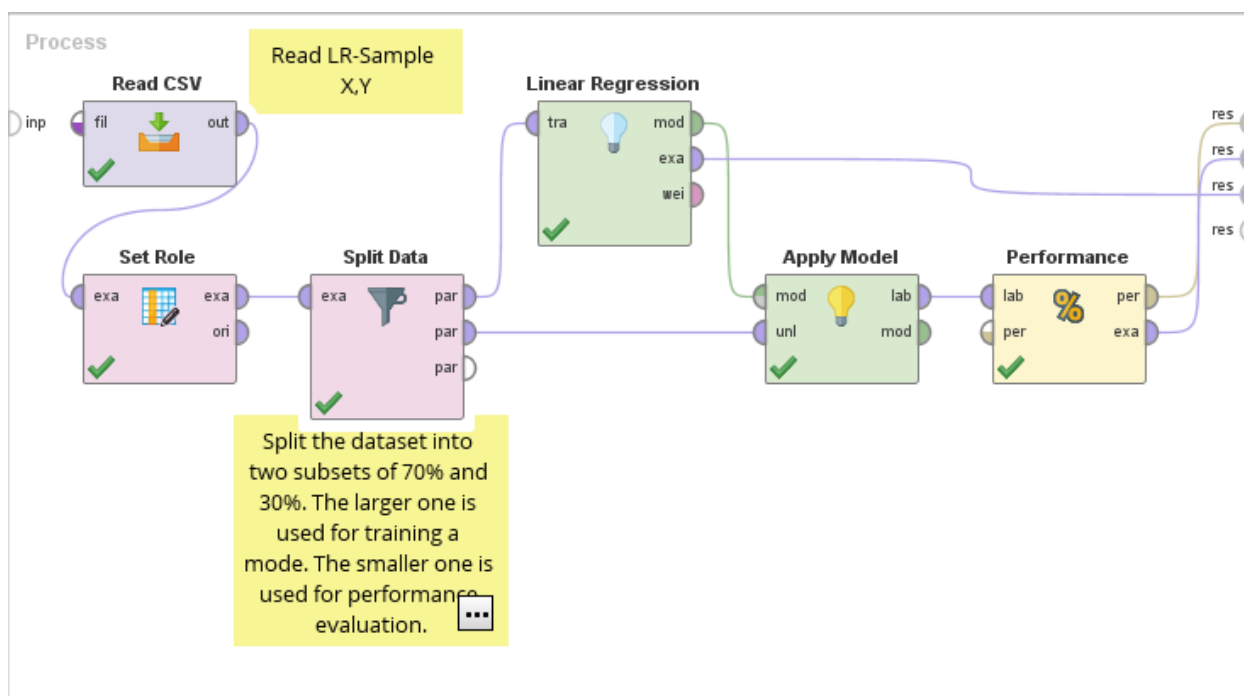
๓.๓ การสร้างและฝึกโมเดล (Modeling & Machine Learning)

โปรแกรม RapidMiner มีโมเดล Machine Learning หลากหลายประเภท เช่น:

- **Supervised Learning** : ใช้สำหรับการทำนาย เช่น Decision Tree, Random Forest, SVM, Deep Learning
- **Unsupervised Learning** : ใช้สำหรับการจำแนกกลุ่ม เช่น K-Means Clustering, DBSCAN, PCA
- **Text Mining** : วิเคราะห์ข้อความ เช่น Sentiment Analysis, Topic Modeling

วิธีการสร้างโมเดล

- ๑) ใช้ Operator Split Data เพื่อแบ่งข้อมูลเป็น Training Set และ Testing Set
- ๒) เลือกโมเดลที่ต้องการ จาก Operators Panel ใช้เชื่อมต่อโมดูล Train Model เพื่อฝึกโมเดลกับข้อมูล Training Set เช่น:
 - Linear Regression สำหรับข้อมูลเชิงปริมาณ
 - Decision Tree หรือ Random Forest สำหรับข้อมูลเชิงคุณภาพ
- ๓) ใช้ Operator Apply Model เพื่อทดสอบโมเดลกับข้อมูล Testing Set
- ๔) ปรับแต่งพารามิเตอร์ของโมเดลด้วย Optimize Parameters
- ๕) กด Run เพื่อฝึกโมเดล



รูปที่ ๓ ตัวอย่างโมเดลในโปรแกรม RapidMiner

๓.๔ การประเมินผลโมเดล (Model Evaluation)

หลังจากฝึกโมเดลแล้ว ต้องมีการวัดความแม่นยำของโมเดลโดยใช้ Performance Metrics เช่น:

- Accuracy : วัดความถูกต้องของโมเดล
- Precision & Recall : ใช้สำหรับปัญหา Classification เช่น การทำนายว่าลูกค้าจะซื้อสินค้าหรือไม่
- RMSE (Root Mean Squared Error) : ใช้สำหรับ Regression เช่น การทำนายยอดขาย

ตัวอย่าง Operators ที่ใช้ในขั้นตอนนี้

- Apply Model : ใช้ทดสอบโมเดลกับข้อมูลใหม่
- Performance (Classification) หรือ Performance (Classification) เพื่อ คำนวณ ประสิทธิภาพ และวัดค่าความแม่นยำของโมเดล
- Cross Validation : ตรวจสอบความสามารถของโมเดลโดยใช้ K-Fold

๓.๕ การนำไปใช้งานจริง (Deployment)

เมื่อโมเดลพร้อมใช้งาน สามารถนำไปใช้ในระบบต่างๆ ได้ เช่น:

- บันทึกเป็นไฟล์โมเดล (PMML, JSON, XML)
- ใช้ Web Service เพื่อสร้าง API นำโมเดลไปใช้งาน
- เชื่อมต่อ RapidMiner Server เพื่อทำงานแบบเรียลไทม์
- เชื่อมต่อกับระบบฐานข้อมูลหรือ Dashboard หรือส่งออกเป็นไฟล์ Excel, CSV

๔. ประโยชน์ที่ได้รับต่อตนเอง:

๑. ทำให้สามารถวิเคราะห์ข้อมูลที่มีเพื่อการตัดสินใจอย่างมีประสิทธิภาพเพิ่มขึ้น
๒. พัฒนาทักษะในการจัดการและวิเคราะห์ข้อมูล ที่สามารถนำมาประยุกต์ใช้กับการทำงานวิจัยได้

๕. ประโยชน์ที่ได้รับต่อหน่วยงาน

องค์ความรู้ที่ได้รับไปประยุกต์ใช้ในการวิเคราะห์ข้อมูลในการศึกษา ทดลอง วิจัยด้านการเกษตร สามารถทำงานได้โดยไม่ต้องมีพื้นฐานการเขียนโปรแกรม มีขั้นตอนที่ง่ายและมีประสิทธิภาพ ตั้งแต่การเก็บข้อมูล การเตรียมข้อมูล การวิเคราะห์ ไปจนถึงการสร้างโมเดล ประเมินผล Machine Learning และการนำไปใช้งานจริง



สำนักงานคณะกรรมการข้าราชการพลเรือน
ขอมอบประกาศนียบัตรฉบับนี้ให้เพื่อแสดงว่า

นางสาวอภันตรี พุดกษพงศ์

ได้ผ่านการพัฒนาทางไกลด้วยระบบอิเล็กทรอนิกส์

วิชา การใช้โปรแกรมดิจิทัลเพื่อการวิเคราะห์ข้อมูล

(รวมระยะเวลาทั้งสิ้น 3 ชั่วโมง)
ให้ไว้ ณ วันที่ 21 กุมภาพันธ์ พ.ศ. 2568

(นายปวิญฉัตร ศิวรักษ์)
เลขาธิการคณะกรรมการข้าราชการพลเรือน

