Machine learning and Deep Learning with KNIME

ในปัจจุบันองค์กรทั้งภาครัฐและเอกชนมีการจัดเก็บข้อมูลเพิ่มขึ้นอย่างต่อเนื่อง และมีความจำเป็นต้องใช้ข้อมูลเพื่อสนับสนุนการตัดสินใจ การคาดการณ์แนวโน้ม และการสร้างความได้เปรียบในการดำเนินงาน เทคโนโลยีด้าน Machine Learning และ Deep Learning จึงเข้ามามีบทบาทสำคัญในการวิเคราะห์ข้อมูลเชิงลึกและสร้างแบบจำลองเชิงคาดการณ์ให้กับองค์กรได้อย่างมีประสิทธิภาพ

อย่างไรก็ตาม ผู้เริ่มต้นจำนวนมากมักเผชิญข้อจำกัดด้านการเขียนโปรแกรมหรือการสร้าง pipeline วิเคราะห์ข้อมูลที่ซับซ้อน เครื่องมือ KNIME Analytics Platform ซึ่งเป็นแพลตฟอร์มแบบ visual workflow และเป็นซอฟต์แวร์ open source จึงเป็นทางเลือกที่เหมาะสมสำหรับการเรียนรู้และประยุกต์ใช้ด้าน Data Analytics, Machine Learning และ Deep Learning เพราะช่วยให้ผู้เรียนมองเห็นลำดับการทำงานของข้อมูลอย่างเป็นระบบ ตั้งแต่การนำเข้าข้อมูล การปรับปรุงคุณภาพข้อมูล การสร้างโมเดล การประเมินผล และการประยุกต์ใช้งานจริง

หลักสูตรนี้จึงถูกออกแบบขึ้นเพื่อให้ผู้เข้าอบรมสามารถเรียนรู้ตั้งแต่พื้นฐานของการเตรียมข้อมูล ไปจนถึงการสร้างโมเดล Machine Learning และ Deep Learning ด้วย KNIME โดยเน้นทั้งความเข้าใจเชิงแนวคิดและการลงมือปฏิบัติจริง ผู้เรียนจะได้ฝึกสร้าง workflow สำหรับงาน classification, association rule, clustering, TensorFlow, deep learning, convolutional neural network และ LSTM

วัตถุประสงค์:

- อธิบายแนวคิดพื้นฐานของ Data Preparation, Machine Learning และ Deep Learning ได้อย่างถูกต้อง
- ใช้งาน KNIME Analytics Platform เพื่อสร้างและจัดการ workflow สำหรับงานวิเคราะห์ข้อมูลได้
- นำเข้าข้อมูลจากแหล่งข้อมูลต่างๆ และดำเนินการ preprocessing
  เช่น type conversion, filtering, join, split, aggregation, pivot, binning และ normalization ได้
- ใช้เครื่องมือ visualization ภายใน KNIME เพื่อสำรวจรูปแบบและคุณลักษณะของข้อมูลเบื้องต้นได้
- สร้างโมเดล Machine Learning สำหรับงาน classification ด้วยอัลกอริทึมสำคัญ
  เช่น Naive Bayes, Decision Tree, k-Nearest Neighbour และ Support Vector Machine ได้
- ประเมินประสิทธิภาพของโมเดลด้วยแนวทางที่เหมาะสม และสามารถเปรียบเทียบผลของอัลกอริทึมแต่ละชนิดได้
- สร้าง workflow สำหรับการทำ Association Rule Mining และ Clustering ด้วยเทคนิค
  เช่น k-Means, Hierarchical Clustering และ DBSCAN ได้
- อธิบายหลักการพื้นฐานของ Deep Learning และเชื่อมการใช้งาน TensorFlow ใน KNIME ได้
- สร้างโมเดลเบื้องต้นสำหรับงาน Deep Learning, CNN และ LSTM ภายใน KNIME ได้
- ประยุกต์ใช้ KNIME เพื่อสร้าง workflow ต้นแบบสำหรับงานวิเคราะห์ข้อมูลหรือการพยากรณ์ในบริบทองค์กรได้

กลุ่มเป้าหมาย:

- Data Engineer
- Data Scientist
- Data Analyst
- นักพัฒนาระบบหรือผู้ปฏิบัติงานด้านข้อมูล
- บุคลากรที่ต้องการเริ่มต้นเรียนรู้ Machine Learning และ Deep Learning ด้วยเครื่องมือแบบ visual workflow
- อาจารย์ วิทยากร หรือผู้สอนที่ต้องการเครื่องมือสำหรับสาธิตกระบวนการวิเคราะห์ข้อมูลและสร้างโมเดลให้เห็นภาพได้ง่าย
- สำหรับความรู้พื้นฐาน หลักสูตรต้นทางระบุว่า ไม่ต้องมีความรู้พื้นฐานใดๆ มาก่อน จึงสามารถใช้เป็นหลักสูตรเริ่มต้นได้ดี โดยเฉพาะสำหรับผู้ที่สนใจงานข้อมูลแต่ไม่ได้ถนัดการเขียนโค้ดมากนัก

ซอฟต์แวร์ที่ใช้:

- **KNIME Analytics Platform** สำหรับสร้าง workflow งาน Data Analytics, Machine Learning และ Deep Learning
- **TensorFlow Integration for KNIME** หรือ extension ที่เกี่ยวข้องสำหรับการทำ Deep Learning ตามแนว course outline ที่ระบุเรื่อง TensorFlow และ Deep Learning
- **ชุดข้อมูลตัวอย่าง** สำหรับงาน classification, clustering, association rule, image classification และ sequence/time-series
- **ระบบปฏิบัติการ** Windows เป็นหลัก หรือสามารถใช้ macOS/Linux ได้หากติดตั้ง KNIME ได้สมบูรณ์
- **Microsoft Excel** หรือโปรแกรมตารางคำนวณ สำหรับตรวจสอบข้อมูลตัวอย่างในบางกิจกรรม
- **เว็บเบราว์เซอร์** สำหรับดาวน์โหลดข้อมูลหรือเอกสารประกอบการอบรม

ข้อกำหนดเครื่องคอมพิวเตอร์เบื้องต้น:

- RAM อย่างน้อย 8 GB
- CPU ระดับ Intel Core i5 / Ryzen 5 หรือสูงกว่า
- พื้นที่ว่างอย่างน้อย 10–20 GB
- อินเทอร์เน็ตสำหรับดาวน์โหลด extensions และ dataset
- แนะนำให้เตรียมเครื่องที่สามารถติดตั้ง KNIME และ TensorFlow extensions ได้เรียบร้อยก่อนเริ่มอบรม

ความรู้พื้นฐาน:

- ไม่ต้องมีความรู้พื้นฐานใดๆ มาก่อน

ระยะเวลาในการอบรม:

- 18 ชั่วโมง (3 วัน)

วิทยากรผู้สอน:

- อาจารย์สามิตร โกยม

เนื้อหาการอบรม:

วันที่ 1: พื้นฐาน KNIME และการเตรียมข้อมูลสำหรับงานวิเคราะห์

ชั่วโมงที่ 1: Introduction to KNIME และภาพรวมการทำงานแบบ Visual Workflow

- ความรู้เบื้องต้นเกี่ยวกับ KNIME
- แนวคิดของ node-based workflow
- ส่วนประกอบของหน้าจอโปรแกรม
- การสร้างและบันทึก workflow
- การ execute node และตรวจสอบผลลัพธ์

ชั่วโมงที่ 2: การนำเข้าข้อมูลและการสำรวจข้อมูลเบื้องต้น

- การอ่านข้อมูลจาก CSV, Excel และ text file
- การตรวจสอบ schema และประเภทข้อมูล
- การสำรวจ dataset เบื้องต้น
- การตรวจสอบ missing values และข้อมูลผิดรูปแบบ

ชั่วโมงที่ 3: Data Manipulation และ Preprocessing I

- Type Conversion
- Formula สำหรับสร้างคอลัมน์ใหม่
- String Replace / Rule-based Replace
- การจัดการค่าว่าง
- การปรับรูปแบบข้อมูลให้พร้อมวิเคราะห์

ชั่วโมงที่ 4: Data Manipulation และ Preprocessing II

- Filtering ข้อมูล
- Column Filter / Row Filter
- การ split และ combine data
- Concatenation
- การตรวจสอบผลลัพธ์หลังแปลงข้อมูล

ชั่วโมงที่ 5: Join, Grouping, Aggregation และ Pivot

- Join ข้อมูลจากหลายแหล่ง
- Grouping and Aggregation
- การสรุปผลเชิงกลุ่ม
- Pivot Operation
- การ reshape ข้อมูลเพื่อใช้งานต่อ

ชั่วโมงที่ 6: Data Transformation เพิ่มเติมและ Visualization เบื้องต้น

- Discrete & Binning
- Normalization
- การใช้ visualization เพื่อดู pattern ของข้อมูล
- Workshop: สร้าง workflow เตรียมข้อมูลครบกระบวนการ

วันที่ 2: Machine Learning ด้วย KNIME

ชั่วโมงที่ 1: Introduction to Machine Learning

- แนวคิดของ Machine Learning
- Supervised และ Unsupervised Learning
- Train/Test Dataset
- การเตรียมข้อมูลก่อนเข้าโมเดล
- แนวคิดเรื่อง data leakage

ชั่วโมงที่ 2: Evaluation of Algorithms

- Confusion Matrix
- Accuracy, Precision, Recall, F1-score
- ROC และ AUC
- การอ่านผลประเมินโมเดล
- การเปรียบเทียบอัลกอริทึมเบื้องต้น

ชั่วโมงที่ 3: Classification Algorithms I

- Naive Bayes
- Decision Tree
- หลักการและ workflow ของแต่ละโมเดล
- การ train และการ prediction

ชั่วโมงที่ 4: Classification Algorithms II

- k-Nearest Neighbour
- Support Vector Machines
- การเลือกอัลกอริทึมให้เหมาะกับลักษณะข้อมูล
- Workshop เปรียบเทียบ classification algorithms

ชั่วโมงที่ 5: Association Rule Mining

- แนวคิด Market Basket Analysis
- Support, Confidence, Lift
- การสร้าง association rule workflow
- การแปลผลกฎความสัมพันธ์

ชั่วโมงที่ 6: Clustering Techniques

- k-Means Clustering
- Optimum Number of K
- Hierarchical Clustering
- DBSCAN
- Comparison of Clustering Algorithms
- Workshop: Customer Segmentation หรือ Data Grouping

วันที่ 3: Deep Learning with KNIME, TensorFlow, CNN และ LSTM

ชั่วโมงที่ 1: Introduction to Deep Learning

- ความแตกต่างระหว่าง Machine Learning และ Deep Learning
- Neural Network เบื้องต้น
- Hidden Layer, Activation Function, Loss Function
- use case ที่เหมาะกับ Deep Learning

ชั่วโมงที่ 2: TensorFlow with KNIME

- ภาพรวม TensorFlow integration
- การเตรียมข้อมูลสำหรับ neural network
- การสร้าง workflow สำหรับ deep learning
- การ train และ monitor model

ชั่วโมงที่ 3: Deep Learning for Classification

- การสร้างโมเดล deep learning สำหรับ classification
- การตั้งค่าพื้นฐานของโมเดล
- การประเมินผล accuracy/loss
- การสังเกตปัญหา overfitting

ชั่วโมงที่ 4: Convolutional Neural Networks และ Computer Vision

- หลักการ CNN
- Convolution, Pooling, Feature Map
- การเตรียมข้อมูลรูปภาพ
- Image Classification เบื้องต้นด้วย KNIME

ชั่วโมงที่ 5: LSTM และ Sequence Modeling

- แนวคิด sequence data
- พื้นฐานของ recurrent model
- LSTM คืออะไร
- ตัวอย่างงาน time series หรือ sequence prediction
- การเตรียมข้อมูลลำดับเวลา

ชั่วโมงที่ 6: Mini Project / Capstone Workshop

- สร้าง workflow แบบครบกระบวนการ
- เลือกโจทย์จาก classification, clustering หรือ deep learning
- สรุปการประยุกต์ใช้ KNIME ในงานจริง
- ถามตอบและแนวทางต่อยอดหลังจบหลักสูตร