สถาบันไอทีจีเนียส
IT GENIUS INSTITUTE

COURSE ID
DBC-02-13

DATA SCIENCE / DATA MINING

COURSE OUTLINE · รายละเอียดหลักสูตรอบรม

Basic Data Mining with R

ค้นหารูปแบบในข้อมูลด้วย R ทั้งจำแนกประเภท จัดกลุ่ม และหากฎความสัมพันธ์

- 18 ชั่วโมง
- 3 วัน
- พื้นฐาน ภาษา R เบื้องต้น
- ระดับ Beginner



ระยะเวลา: 3 วัน
18 ชั่วโมง (3 วัน วันละ 6 ชม.) · 09:00-16:00 น.

รูปแบบ: Classroom / Hybrid
Onsite หรือ Online ผ่าน Microsoft Teams · Public / In-house

รายละเอียดหลักสูตร
www.itgenius.co.th

ข้อมูลที่องค์กรและนักวิจัยเก็บไว้ ทั้งข้อมูลลูกค้า ยอดขาย แบบสอบถาม และผลการเรียน มีรูปแบบซ่อนอยู่มากมาย Data Mining คือการค้นหารูปแบบเหล่านั้นเพื่อใช้พยากรณ์ จัดกลุ่ม หรือหาความสัมพันธ์ ทั้งในงานวิทยานิพนธ์ งานวิจัย และการตลาด ภาษา R เป็นเครื่องมือโอเพนซอร์สที่ใช้ฟรีและนิยมในงานสถิติและวิจัย แต่ผู้เริ่มต้นมักติดอยู่ที่การเตรียมข้อมูล การเลือกเทคนิค และการวัดผลแบบจำลองให้ถูกต้อง

หลักสูตรนี้ต่อยอดจากพื้นฐานภาษา R พาผู้เรียนทำ Data Mining ตามกระบวนการ CRISP-DM ด้วย R 4.6 และชุด Package รุ่นปัจจุบัน เริ่มจากนำเข้า สำรวจ และทำความสะอาดข้อมูล เตรียมตัวแปรด้วย recipes จากนั้นสร้างแบบจำลองจำแนกประเภทด้วย tidymodels ทั้ง Decision Tree, K-Nearest Neighbors, Logistic Regression, Neural Network และ Random Forest พร้อมวัดผลด้วย Cross-validation และปรับจูน ต่อด้วยการจัดกลุ่มด้วย k-means และ Hierarchical Clustering การหากฎความสัมพันธ์ด้วย arules และรายงานผลด้วย Quarto ก่อนปิดท้ายด้วย Workshop โปรเจกต์ตั้งแต่ต้นจนจบ (อบรม 3 วัน วันละ 6 ชั่วโมง รวม 18 ชั่วโมง ระดับ Beginner)

เครื่องมือ
- R 4.6
- tidymodels
- Decision Tree
- arules
- Quarto

ราคาค่าอบรม · PUBLIC / IN-HOUSE

6,500 บาท / ท่าน

สมัครผ่านเว็บไซต์ ลด 10%

ราคายังไม่รวมภาษีมูลค่าเพิ่ม (VAT 7%) · รวมเอกสารประกอบการอบรม Lab Guide, ชุดข้อมูลและสคริปต์ R ครบทุก Lab, Template รายงาน Quarto และ Certificate · รับจัด In-house ขอใบเสนอราคาได้

เพิ่มเพื่อน LINE

02-570-8449 · 088-807-9770 · @itgenius · www.itgenius.co.th

01 วัตถุประสงค์ Objectives

- เข้าใจหลักการ Data Mining ประเภทของงาน และขั้นตอนตามกระบวนการ CRISP-DM
- นำเข้า สำรวจ ทำความสะอาด และเตรียมข้อมูลให้พร้อมวิเคราะห์ด้วย R
- สร้างแบบจำลองจำแนกประเภทด้วย Decision Tree, K-Nearest Neighbors และ Neural Network ผ่าน tidymodels
- วัดผลแบบจำลองด้วย Confusion Matrix, ROC AUC และ Cross-validation แล้วเลือกแบบจำลองที่เหมาะสม
- จัดกลุ่มข้อมูลด้วย Clustering และหากฎความสัมพันธ์ด้วย Association Rules
- นำผลไปใช้ในงานวิทยานิพนธ์ งานวิจัย หรือการตลาด และเขียนรายงานที่ทำซ้ำได้ด้วย Quarto

02 กลุ่มเป้าหมาย Target Audience

- นักศึกษาปริญญาโทและเอกที่ต้องใช้ Data Mining ในวิทยานิพนธ์
- นักวิจัยและอาจารย์ที่ต้องวิเคราะห์ข้อมูลงานวิจัยด้วย R
- นักการตลาดและนักวิเคราะห์ข้อมูลที่ต้องแบ่งกลุ่มลูกค้าหรือพยากรณ์พฤติกรรม
- เจ้าหน้าที่หน่วยงานที่ดูแลข้อมูลและต้องการต่อยอดจาก R พื้นฐาน

03 จุดเด่นของหลักสูตร Highlights

- **ต่อยอดจาก R พื้นฐาน** R 4.6, RStudio หรือ Positron และ tidyverse
- **ใช้เครื่องมือรุ่นปัจจุบัน** tidymodels, recipes และ yardstick
- **ครบสามกลุ่มเทคนิค** Classification, Clustering, Association
- **วัดผลอย่างถูกวิธี** Train/Test, Cross-validation และ ROC
- **พร้อมใช้ในงานวิจัย** รายงานที่ทำซ้ำได้ด้วย Quarto

04 ความรู้พื้นฐาน Prerequisites

- เขียนภาษา R เบื้องต้นได้ เช่น ตัวแปร Data Frame และการใช้ Package หรือผ่านหลักสูตร Basic R Programming
- มีความรู้สถิติพื้นฐาน เช่น ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และสหสัมพันธ์
- ใช้งานคอมพิวเตอร์และจัดการไฟล์ได้คล่อง
- Notebook ที่ติดตั้ง R, RStudio หรือ Positron และ Package ได้ พร้อมสิทธิ์ติดตั้งโปรแกรม

05 สิ่งที่ต้องเตรียมและเครื่องมือ Setup & Tools

รายการ	รายละเอียด
อุปกรณ์	Notebook Windows หรือ macOS เชื่อมต่ออินเทอร์เน็ต หน่วยความจำ 8 GB ขึ้นไป
ภาษาและ IDE	R 4.6 กับ RStudio หรือ Positron และ Quarto
Package หลัก	tidyverse, tidymodels, rpart.plot, kknn, ranger, tidyclust และ cluster
Association Rules	arules และ arulesViz
ไฟล์ฝึก	ชุดข้อมูลลูกค้า ยอดขาย ตะกร้าสินค้า และแบบสอบถาม พร้อมสคริปต์เฉลย

หมายเหตุ: ปรับปรุงใหม่ปี 2026 จากหลักสูตรเดิม: ใช้ R 4.6 กับ RStudio หรือ Positron และเปลี่ยนมาสร้างแบบจำลองด้วยชุด tidymodels ซึ่งเป็นแนวทางปัจจุบันของ R ทั้ง recipes, parsnip, workflows, rsample, tune และ yardstick แทนการเรียกฟังก์ชันของแต่ละ Package แยกกัน เพิ่มกระบวนการ CRISP-DM การแบ่งข้อมูล Train และ Test, Cross-validation, การปรับจูน, Random Forest, การวัดคุณภาพของกลุ่มด้วย Silhouette, การรายงานด้วย Quarto และจริยธรรมข้อมูลตาม PDPA เครื่องมือทั้งหมดเป็นโอเพนซอร์สและใช้ฟรี หลักสูตรนี้ต่อยอดจากหลักสูตร Basic R Programming ซึ่งสอนพื้นฐานภาษา R การจัดการข้อมูลและกราฟ จึงทบทวน R เพียงสั้น ๆ ในวันแรก

06 ภาพรวมการเรียน 3 วัน Big Picture

วัน	หัวข้อหลัก
Day 1 Section 1-5	รู้จัก Data Mining และกระบวนการ CRISP-DM, ทบทวน R และเตรียมโปรเจกต์, นำเข้าและสำรวจข้อมูล, ทำความสะอาดข้อมูล และเตรียมตัวแปรด้วย recipes พร้อมแบ่งข้อมูล Train และ Test
Day 2 Section 6-10	จำแนกประเภทด้วย Decision Tree, วัดผลด้วย Confusion Matrix, ROC และ Cross-validation, K-Nearest Neighbors และ Logistic Regression, Neural Network และ Random Forest แล้วปรับจูนและเลือกแบบจำลอง
Day 3 Section 11-15	จัดกลุ่มข้อมูลด้วย k-means และ Hierarchical Clustering, หากฎความสัมพันธ์ด้วย Apriori, รายงานผลด้วย Quarto และ Workshop โปรเจกต์ Data Mining ตั้งแต่ต้นจนจบ

DAY 1

กระบวนการ Data Mining และการเตรียมข้อมูล

เนื้อหาการอบรม · Section 1-5

รวม
6 ชั่วโมง

1 Data Mining และกระบวนการ CRISP-DM

- Data Mining คืออะไร และต่างจากสถิติเชิงพรรณนาอย่างไร
- งานแบบมีผู้สอน ไม่มีผู้สอน และการหากฎความสัมพันธ์
- ขั้นตอน CRISP-DM ตั้งแต่เข้าใจโจทย์จนนำผลไปใช้
- ตัวอย่างในงานวิทยานิพนธ์ งานวิจัย และการตลาด
- จริยธรรมข้อมูล ความเป็นส่วนตัว และ PDPA

2 Lab: ทบทวน R และเตรียมโปรเจกต์

- ทบทวน R 4.6 กับ RStudio หรือ Positron และการสร้าง Project
- ติดตั้ง tidymodels และ Package ที่ใช้ในหลักสูตร และล็อกรุ่นด้วย renv
- ทบทวน dplyr และ Native Pipe สำหรับจัดการข้อมูล
- ตั้งค่า set.seed() เพื่อให้ผลการวิเคราะห์ทำซ้ำได้

3 Lab: นำเข้าและสำรวจข้อมูล

- นำเข้าข้อมูลจาก CSV, Excel และฐานข้อมูลผ่าน DBI
- สรุปภาพรวมของตัวแปรและชนิดข้อมูล
- สำรวจการแจกแจงและความสัมพันธ์ด้วย ggplot2
- หาสหสัมพันธ์ระหว่างตัวแปรเชิงตัวเลข
- Lab: สำรวจข้อมูลลูกค้าและตั้งคำถามที่ต้องการตอบ

4 Lab: ทำความสะอาดข้อมูล

- จัดการค่าที่หายไป ค่าซ้ำ และค่าผิดปกติ
- แปลงชนิดข้อมูลและจัดการตัวแปรเชิงกลุ่มด้วย factor
- ข้อความภาษาไทยและ Encoding แบบ UTF-8
- Lab: ทำความสะอาดข้อมูลแบบสอบถามให้พร้อมวิเคราะห์

5 Lab: เตรียมตัวแปรด้วย recipes

- แบ่งข้อมูล Train และ Test ด้วย initial_split แบบ Stratified
- ปรับสเกลตัวเลข สร้าง Dummy Variable และเติมค่าที่หายไปใน recipe
- รวมกลุ่มที่มีจำนวนน้อยและตัดตัวแปรที่ไม่มีประโยชน์
- Data Leakage คืออะไร และทำไมต้องเตรียมข้อมูลจากชุด Train เท่านั้น
- Lab: สร้าง recipe ที่ใช้ซ้ำได้กับทุกแบบจำลอง

DAY 2

การจำแนกประเภทข้อมูล (Classification)

เนื้อหาการอบรม · Section 6-10

รวม
6 ชั่วโมง

6 Lab: Decision Tree

- หลักการของ Decision Tree และการแบ่งกิ่งด้วย Gini
- สร้างแบบจำลองด้วย decision_tree() และ Engine rpart ใน workflow
- วาดและอ่านต้นไม้ด้วย rpart.plot แล้วแปลงเป็นกฎที่อธิบายได้
- Cost Complexity และความลึกของต้นไม้กับปัญหา Overfitting
- Lab: พยากรณ์ลูกค้าที่มีแนวโน้มเลิกใช้บริการ

7 Lab: วัดผลแบบจำลอง

- Confusion Matrix, Accuracy, Precision, Recall และ F1
- ROC Curve และ ROC AUC ด้วย yardstick
- Cross-validation ด้วย vfold_cv และ fit_resamples
- ข้อมูลไม่สมดุลและเหตุที่ Accuracy อย่างเดียวไม่พอ
- Lab: เปรียบเทียบผลของ Decision Tree บนชุด Train และ Test

8 Lab: K-Nearest Neighbors และ Logistic Regression

- หลักการของ K-Nearest Neighbors และการเลือกค่า k
- ทำไม KNN ต้องปรับสเกลข้อมูลก่อน
- Logistic Regression เป็นแบบจำลองฐานสำหรับเปรียบเทียบ
- อ่านค่าสัมประสิทธิ์และ Odds Ratio เพื่ออธิบายผล
- Lab: จำแนกกลุ่มลูกค้าด้วย KNN และ Logistic Regression

9 Lab: Neural Network และ Random Forest

- หลักการของ Neural Network แบบชั้นซ่อนชั้นเดียว
- สร้างแบบจำลองด้วย mlp() และ Engine nnet
- Random Forest ด้วย Engine ranger และแนวคิด Ensemble
- ความสำคัญของตัวแปร (Variable Importance)
- Lab: เปรียบเทียบความแม่นยำกับแบบจำลองก่อนหน้า

10 Lab: ปรับจูนและเลือกแบบจำลอง

- ปรับจูน Hyperparameter ด้วย tune_grid()
- เปรียบเทียบหลายแบบจำลองพร้อมกันด้วย workflow_set
- เลือกแบบจำลองที่ดีที่สุดและประเมินครั้งสุดท้ายด้วย last_fit()
- บันทึกแบบจำลองและพยากรณ์ข้อมูลใหม่
- Lab: เลือกแบบจำลองที่เหมาะสมที่สุดสำหรับโจทย์ลูกค้า

DAY 3

Clustering, Association Rules และการรายงานผล

เนื้อหาการอบรม · Section 11-15

รวม
6 ชั่วโมง

11 Lab: จัดกลุ่มข้อมูลด้วย k-means

- หลักการของ Clustering และการวัดระยะห่าง
- ปรับสเกลข้อมูลก่อนจัดกลุ่ม
- จัดกลุ่มด้วย kmeans() และ k_means() ใน tidyclust
- เลือกจำนวนกลุ่มด้วย Elbow Method และ Silhouette
- Lab: แบ่งกลุ่มลูกค้าตามพฤติกรรมการซื้อ

12 Lab: Hierarchical Clustering และการแปลผลกลุ่ม

- Hierarchical Clustering และการอ่าน Dendrogram
- วิธีเชื่อมกลุ่มแบบ Complete, Average และ Ward
- สรุปลักษณะของแต่ละกลุ่มด้วยตารางและกราฟ
- ลดมิติด้วย PCA เพื่อแสดงกลุ่มบนกราฟสองมิติ
- Lab: ตั้งชื่อและอธิบายกลุ่มลูกค้าให้ใช้วางแผนการตลาดได้

13 Lab: หากฎความสัมพันธ์ (Association Rules)

- Market Basket Analysis และข้อมูลแบบ Transaction
- Support, Confidence และ Lift
- หากฎด้วย apriori() ใน arules และคัดกฎด้วยเงื่อนไข
- แสดงกฎด้วย arulesViz
- Lab: หาสินค้าที่ลูกค้ามักซื้อคู่กันจากข้อมูลตะกร้าสินค้า

14 Lab: รายงานผลด้วย Quarto

- โครงรายงาน Data Mining สำหรับวิทยานิพนธ์และงานวิจัย
- สร้างรายงาน Quarto ที่รวมโค้ด ตาราง กราฟ และคำอธิบาย
- อธิบายผลแบบจำลองให้ผู้อ่านที่ไม่ใช่สายเทคนิคเข้าใจ
- ใช้ AI Assistant ช่วยอธิบายโค้ดและตรวจผลอย่างระมัดระวัง

15 Workshop: โปรเจกต์ Data Mining ตั้งแต่ต้นจนจบ

- เลือกโจทย์และชุดข้อมูลตามสายงานหรืองานวิจัยของตนเอง
- เตรียมข้อมูล สร้างและเปรียบเทียบแบบจำลองตามกระบวนการ CRISP-DM
- สรุปผลเป็นรายงาน Quarto ที่ทำซ้ำได้
- นำเสนอผลงานและรับคำแนะนำจากวิทยากร

07 เรียนจบแล้วได้อะไร Outcomes

- **เข้าใจกระบวนการ** Data Mining และ CRISP-DM
- **เตรียมข้อมูลเป็น** ทำความสะอาดและ recipes
- **สร้างแบบจำลองจำแนกได้** Decision Tree, KNN, Neural Network
- **วัดผลอย่างถูกวิธี** Cross-validation และ ROC AUC
- **จัดกลุ่มและหากฎได้** k-means, Hierarchical, Apriori
- **รายงานผลได้** Quarto สำหรับงานวิจัยและธุรกิจ

08 สิ่งที่ผู้เรียนจะได้รับ Deliverables

- เอกสารประกอบการอบรมครบทั้ง 3 วัน และ Lab Guide ทีละขั้นตอน
- ชุดข้อมูลตัวอย่างและสคริปต์ R เฉลยครบทุก Lab
- Template รายงาน Data Mining ด้วย Quarto สำหรับงานวิจัยและงานธุรกิจ
- ประกาศนียบัตรจบหลักสูตร (Certificate of Completion) และช่องทางสอบถามวิทยากรหลังจบหลักสูตร

09 ข้อมูลผู้สอน Instructor

อ.สนิทวงศ์ กมลภากรณ์ ผู้สอนที่ใช้ภาษา R วิเคราะห์ข้อมูลในงานวิจัยและงานองค์กร และให้คำปรึกษาการใช้ Data Mining ในวิทยานิพนธ์ เน้นอธิบายหลักการของแต่ละเทคนิคด้วยภาษาเข้าใจง่ายก่อนลงมือเขียนโค้ด ให้ผู้เรียนฝึกกับข้อมูลใกล้ตัว และฝึกอ่านและอธิบายผลให้ถูกต้อง เพื่อให้กลับไปเลือกเทคนิคและวิเคราะห์ข้อมูลของตนเองได้อย่างมั่นใจ

ราคาค่าอบรม 6,500 บาท/ท่าน · สนใจ IN-HOUSE ขอใบเสนอราคา

6,500 บาท / ท่าน (ยังไม่รวม VAT 7%)

สมัครผ่านเว็บไซต์ ลด 10%

02-570-8449 · 088-807-9770

@itgenius · contact@itgenius.co.th

เพิ่มเพื่อน LINE