สถาบันไอทีจีเนียส
IT GENIUS INSTITUTE

COURSE ID
WDC-37
DATA SCIENCE / PYTHON

COURSE OUTLINE · รายละเอียดหลักสูตรอบรม

Data Science with Python

สร้างแบบจำลองพยากรณ์จากข้อมูลจริง ด้วย Python ตามกระบวนการ CRISP-DM

- 18 ชั่วโมง
- 3 วัน
- พื้นฐาน Python เบื้องต้น
- ระดับ Intermediate



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

องค์กรส่วนใหญ่มีข้อมูลยอดขาย ลูกค้า และการใช้งานระบบสะสมอยู่มาก แต่ยังใช้ข้อมูลเหล่านั้นเพียงทำรายงานย้อนหลัง ทั้งที่ข้อมูลชุดเดียวกันนี้สามารถนำมาสร้างแบบจำลองพยากรณ์ได้ เช่น คาดการณ์ยอดขาย ทำนายลูกค้าที่มีแนวโน้มยกเลิกบริการ แบ่งกลุ่มลูกค้า หรือหาสินค้าที่มักถูกซื้อคู่กัน ภาษา Python มีเครื่องมือสำหรับงานเหล่านี้ครบชุด และกลายเป็นภาษามาตรฐานของงาน Data Science ในปัจจุบัน

หลักสูตรนี้พาผู้เรียนทำ Predictive Analytics ตามกระบวนการ CRISP-DM ตั้งแต่ตั้งโจทย์ธุรกิจ เตรียมและทำความสะอาดข้อมูลด้วย pandas 3 สำรวจข้อมูลด้วยกราฟและสถิติ สร้าง Feature และ Pipeline ด้วย scikit-learn แล้วสร้างแบบจำลองด้วย Linear และ Logistic Regression, K-Means Clustering, Decision Tree และ Ensemble, Association Rules และ Neural Network เบื้องต้น พร้อมวัดผลอย่างถูกวิธี และปิดท้ายด้วยการนำเสนอผลและโปรเจกต์พยากรณ์จากข้อมูลจริง (อบรม 3 วัน วันละ 6 ชั่วโมง รวม 18 ชั่วโมง ระดับ Intermediate)

เครื่องมือ: Python · pandas 3 · scikit-learn · JupyterLab · CRISP-DM

ราคาค่าอบรม · PUBLIC / IN-HOUSE

9,500 บาท / ท่าน

สมัครผ่านเว็บไซต์ ลด 10%

ราคายังไม่รวมภาษีมูลค่าเพิ่ม (VAT 7%) · รวมเอกสารประกอบการอบรม Notebook และชุดข้อมูลตัวอย่างครบทุก Lab และ Certificate · รับจัด In-house ขอใบเสนอราคาได้

เพิ่มเพื่อน LINE

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

01 วัตถุประสงค์ Objectives

- วางโจทย์และขั้นตอนงาน Data Science ตามกระบวนการ CRISP-DM ได้
- เตรียม ทำความสะอาด และแปลงข้อมูลด้วย pandas 3 ได้อย่างถูกต้อง
- สำรวจข้อมูลด้วยกราฟและสถิติเพื่อหาความสัมพันธ์ก่อนสร้างโมเดล
- สร้าง Feature และ Pipeline ด้วย scikit-learn โดยไม่ให้ข้อมูลรั่ว
- สร้างและวัดผลแบบจำลอง Regression และ Classification ด้วยตัวชี้วัดที่เหมาะสม
- แบ่งกลุ่มข้อมูลด้วย K-Means และหากฎความสัมพันธ์ด้วย Association Rules
- ใช้ Decision Tree, Random Forest, Gradient Boosting และ Neural Network เบื้องต้นได้
- สื่อสารผลการวิเคราะห์และบันทึกโมเดลเพื่อนำไปใช้งานต่อ

02 กลุ่มเป้าหมาย Target Audience

- นักวิเคราะห์ข้อมูลและผู้ทำรายงานที่ต้องการก้าวจากรายงานย้อนหลังไปสู่การพยากรณ์
- โปรแกรมเมอร์และฝ่ายไอทีที่ต้องการเริ่มงาน Data Science ด้วย Python
- ทีมการตลาด ขาย และวางแผนที่ต้องใช้ข้อมูลประกอบการตัดสินใจ
- อาจารย์และนักศึกษาที่ต้องการแนวทางทำโปรเจกต์ Predictive Analytics
- ผู้ที่ผ่านหลักสูตร Python เบื้องต้นมาแล้วและต้องการต่อยอดสู่งานข้อมูล

03 จุดเด่นของหลักสูตร Highlights

- **ทำงานตาม CRISP-DM** จากโจทย์ธุรกิจถึงโมเดลที่ใช้งานได้
- **เครื่องมือรุ่นปัจจุบัน** pandas 3, scikit-learn 1.9 และ uv
- **ครบเทคนิคพยากรณ์หลัก** Regression, Clustering, Tree และ Rules
- **วัดผลอย่างถูกวิธี** Pipeline, Cross-Validation และ Metric
- **ลงมือทุกหัวข้อ** Lab บน Notebook และ Capstone Project

04 ความรู้พื้นฐาน Prerequisites

- เขียน Python ระดับพื้นฐานได้ เช่น ตัวแปร List, Dictionary, Loop และ Function
- เข้าใจสถิติเบื้องต้น เช่น ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน (ทบทวนให้ในหลักสูตร)
- ใช้คอมพิวเตอร์จัดการไฟล์และโฟลเดอร์ได้คล่อง
- โน้ตบุ๊กที่มีสิทธิ์ติดตั้งโปรแกรม หรือบัญชี Google สำหรับใช้ Google Colab

05 สิ่งที่ต้องเตรียมและเครื่องมือ Setup & Tools

รายการ	รายละเอียด
อุปกรณ์	Notebook เชื่อมต่ออินเทอร์เน็ต RAM 8 GB ขึ้นไป ไม่ต้องใช้ GPU
ภาษาและเครื่องมือ	Python 3.12 ขึ้นไป, uv, JupyterLab หรือ VS Code และ Google Colab
ไลบรารี	NumPy, pandas 3, Matplotlib, seaborn, scikit-learn 1.9, mlxtend และ skops
ข้อมูลตัวอย่าง	ยอดขายร้านค้า, ลูกค้ายกเลิกบริการ, ตะกร้าสินค้า และข้อมูลสาธารณะ

หมายเหตุ: ปรับปรุงใหม่ปี 2026: ยังคงแกนเดิมคือ CRISP-DM, Data Cleaning, Regression, Clustering, Decision Tree, Association Rules และ Neural Network แต่ย้ายมาใช้ pandas 3 (Copy-on-Write และชนิดข้อมูล str เป็นค่าเริ่มต้น) และ scikit-learn 1.9 จัดการโปรเจกต์ด้วย uv และเพิ่ม Pipeline, Cross-Validation, Gradient Boosting และการนำเสนอผล ส่วนงาน Association Rules ใช้ mlxtend เพราะ scikit-learn ไม่มีให้ หลักสูตรนี้เน้นงานพยากรณ์เชิงธุรกิจ หากต้องการเจาะอัลกอริทึม ML เช่น SVM และ Eigenfaces ดูหลักสูตร Python for Machine Learning เครื่องมือทั้งหมดเป็นโอเพนซอร์ส

06 ภาพรวมการเรียน 3 วัน Big Picture

วัน	หัวข้อหลัก
Day 1 Section 1-5	วางโจทย์ด้วย CRISP-DM, เตรียมเครื่องมือด้วย uv และ JupyterLab, จัดการข้อมูลด้วย pandas 3, ทำความสะอาดข้อมูล, สำรวจข้อมูลด้วยกราฟ และสถิติที่ใช้ในงานพยากรณ์
Day 2 Section 6-10	สร้าง Feature และ Pipeline ที่ไม่รั่วข้อมูล, Linear Regression, Logistic Regression พร้อมการวัดผล, Cross-Validation และการจัดกลุ่มลูกค้าด้วย K-Means
Day 3 Section 11-15	Decision Tree, Random Forest และ Gradient Boosting, หากฎความสัมพันธ์ของสินค้าด้วย Association Rules, Neural Network เบื้องต้น, นำเสนอผลและนำโมเดลไปใช้ และ Capstone Project

DAY 1

เตรียมข้อมูลและเข้าใจข้อมูล

เนื้อหาการอบรม · Section 1-5

รวม
6 ชั่วโมง

1 Lab: Predictive Modelling และ CRISP-DM

- Descriptive, Predictive และ Prescriptive Analytics ต่างกันอย่างไร
- 6 ขั้นตอนของ CRISP-DM จาก Business Understanding ถึง Deployment
- แปลงคำถามธุรกิจเป็นโจทย์ Regression, Classification หรือ Clustering
- ติดตั้งโปรเจกต์ด้วย uv และใช้ JupyterLab, VS Code หรือ Google Colab
- Lab: วางโจทย์พยากรณ์ของร้านค้าตัวอย่างตามกรอบ CRISP-DM

2 Lab: จัดการข้อมูลด้วย pandas 3

- DataFrame, Series และการอ่านข้อมูลจาก CSV, Excel และฐานข้อมูล
- เลือก กรอง เรียง และสร้างคอลัมน์ใหม่ด้วย loc และ assign
- Copy-on-Write และชนิดข้อมูล str ที่เป็นค่าเริ่มต้นใน pandas 3
- groupby, pivot_table และ merge สำหรับสรุปและรวมข้อมูล
- Lab: สรุปยอดขายรายเดือนแยกตามสาขาและหมวดสินค้า

3 Lab: Data Cleaning

- ตรวจคุณภาพข้อมูล: ค่าว่าง ค่าซ้ำ ชนิดข้อมูลผิด และค่าผิดปกติ
- เลือกวิธีจัดการค่าว่าง: ตัดทิ้ง เติมค่า หรือสร้างตัวบ่งชี้
- แปลงวันที่ ข้อความ และหมวดหมู่ให้พร้อมใช้งาน
- หาและจัดการ Outlier ด้วย IQR และ Z-score
- Lab: ทำความสะอาดข้อมูลลูกค้าจริงที่มีปัญหาหลายจุด

4 Lab: Exploratory Data Analysis

- Histogram, Box Plot และ Scatter Plot ด้วย Matplotlib และ seaborn
- ดูการกระจายตัว ความเบ้ และความสัมพันธ์ระหว่างตัวแปร
- Correlation Matrix และข้อควรระวังเรื่องความสัมพันธ์กับเหตุและผล
- สำรวจข้อมูลแยกกลุ่มเพื่อหาสัญญาณที่ใช้พยากรณ์ได้
- Lab: EDA ข้อมูลลูกค้ายกเลิกบริการพร้อมสรุปข้อค้นพบ

5 สถิติสำหรับ Predictive Modelling

- ประชากร ตัวอย่าง และการสุ่มแบ่งข้อมูล Train และ Test
- การแจกแจงปกติ ค่ากลาง และการวัดการกระจาย
- การทดสอบสมมติฐานและ p-value สำหรับคนทำงานข้อมูล
- Bias, Variance, Overfitting และ Underfitting
- Lab: เปรียบเทียบยอดขายสองกลุ่มด้วยสถิติอย่างถูกวิธี

DAY 2

Regression, Classification และ Clustering

เนื้อหาการอบรม · Section 6-10

รวม
6 ชั่วโมง

6 Lab: Feature Engineering และ Pipeline

- เข้ารหัสข้อมูลหมวดหมู่ด้วย OneHotEncoder และ OrdinalEncoder
- ปรับสเกลข้อมูลด้วย StandardScaler และ MinMaxScaler
- สร้าง Feature จากวันที่ ยอดสะสม และพฤติกรรมลูกค้า
- รวมขั้นตอนด้วย Pipeline และ ColumnTransformer
- Data Leakage คืออะไร และป้องกันอย่างไร

7 Lab: Linear Regression

- แนวคิด Linear Regression และการตีความค่าสัมประสิทธิ์
- Multiple Regression และปัญหา Multicollinearity
- วัดผลด้วย MAE, RMSE และ R^2
- Regularization ด้วย Ridge และ Lasso
- Lab: พยากรณ์ยอดขายรายสัปดาห์จากโปรโมชันและฤดูกาล

8 Lab: Logistic Regression และการวัดผล

- Logistic Regression และความน่าจะเป็นของแต่ละกลุ่ม
- Confusion Matrix, Precision, Recall และ F1-score
- ROC Curve, AUC และการเลือก Threshold ตามต้นทุนธุรกิจ
- รับมือข้อมูลไม่สมดุลด้วย class_weight
- Lab: ทำนายลูกค้าที่มีแนวโน้มยกเลิกบริการ

9 Cross-Validation และการปรับ Hyperparameter

- ทำไมการแบ่ง Train และ Test ครั้งเดียวไม่พอ
- K-Fold, Stratified K-Fold และ TimeSeriesSplit
- ค้นหา Hyperparameter ด้วย GridSearchCV และ RandomizedSearchCV
- เปรียบเทียบโมเดลหลายตัวอย่างเป็นธรรม
- Lab: เลือกโมเดลที่ดีที่สุดสำหรับโจทย์ลูกค้ายกเลิกบริการ

10 Lab: Data Clustering

- แนวคิด Unsupervised Learning และการวัดระยะห่าง
- K-Means, Elbow Method และ Silhouette Score
- Hierarchical Clustering และ Dendrogram
- สร้าง Feature แบบ RFM สำหรับแบ่งกลุ่มลูกค้า
- Lab: แบ่งกลุ่มลูกค้าและตั้งชื่อแต่ละกลุ่มเพื่อใช้ทำการตลาด

DAY 3

Tree, Association Rules และการนำไปใช้

เนื้อหาการอบรม · Section 11-15

รวม
6 ชั่วโมง

11 Lab: Decision Tree และ Ensemble

- Decision Tree สร้างกฎการตัดสินใจอย่างไร
- ควบคุมความลึกของต้นไม้และป้องกัน Overfitting
- Random Forest และ HistGradientBoosting
- Feature Importance และ Permutation Importance
- Lab: อธิบายปัจจัยที่ทำให้ลูกค้ายกเลิกบริการด้วยแผนภาพต้นไม้

12 Lab: Association Rules Discovery

- Market Basket Analysis และการเตรียมข้อมูลตะกร้าสินค้า
- Support, Confidence และ Lift
- หา Frequent Itemsets ด้วย Apriori และ FP-Growth ใน mlxtend
- คัดกฎที่มีความหมายทางธุรกิจและตัดกฎที่ไม่มีประโยชน์
- Lab: หาสินค้าที่ควรจัดโปรโมชันคู่กันจากข้อมูลการซื้อ

13 Neural Network เบื้องต้น

- Neuron, Layer, Activation Function และการเรียนรู้แบบ Backpropagation
- สร้าง Neural Network ด้วย MLPClassifier และ MLPRegressor
- เมื่อไร Neural Network คุ้มค่า และเมื่อไรโมเดลแบบดั้งเดิมดีกว่า
- Lab: เปรียบเทียบ Neural Network กับโมเดลที่สร้างมาก่อนหน้า

14 Lab: สื่อสารผลและนำโมเดลไปใช้

- เล่าผลการวิเคราะห์ด้วยกราฟที่ผู้บริหารอ่านเข้าใจ
- บันทึกโมเดลอย่างปลอดภัยด้วย skops และโหลดกลับมาใช้
- ใช้โมเดลพยากรณ์ข้อมูลชุดใหม่และส่งผลออกเป็นไฟล์
- ใช้ AI Assistant ช่วยเขียนโค้ด pandas และตรวจผลก่อนใช้จริง
- ติดตามความแม่นยำเมื่อข้อมูลเปลี่ยนไป (Data Drift)

15 Workshop: Capstone Project

- เลือกโจทย์พยากรณ์จากชุดข้อมูลที่เตรียมไว้หรือข้อมูลของตนเอง
- ทำครบวงจร CRISP-DM ตั้งแต่เตรียมข้อมูลจนวัดผลโมเดล
- นำเสนอผลลัพธ์ ข้อจำกัด และข้อเสนอแนะต่อธุรกิจ
- รีวิวงานร่วมกันและสรุปแนวทางต่อยอด

07 เรียนจบแล้วได้อะไร Outcomes

- **วางโจทย์ได้เป็นระบบ** ทำงานตามขั้นตอน CRISP-DM
- **เตรียมข้อมูลได้คล่อง** Cleaning และ EDA ด้วย pandas 3
- **สร้างโมเดลพยากรณ์** Regression และ Classification
- **แบ่งกลุ่มและหากฎ** K-Means และ Association Rules
- **วัดผลอย่างน่าเชื่อถือ** Pipeline และ Cross-Validation
- **สื่อสารและนำไปใช้** นำเสนอผลและบันทึกโมเดล

08 สิ่งที่ผู้เรียนจะได้รับ Deliverables

- เอกสารประกอบการอบรมฉบับสมบูรณ์ครบทั้ง 3 วัน
- Notebook ตัวอย่างและเฉลยครบทุก Lab พร้อมไฟล์โปรเจกต์ uv
- ชุดข้อมูลตัวอย่างสำหรับฝึกซ้ำหลังจบหลักสูตร
- ประกาศนียบัตรจบหลักสูตร (Certificate of Completion) และช่องทางสอบถามวิทยากรหลังจบหลักสูตร

09 ข้อมูลผู้สอน Instructor

อ.สนิทวงศ์ กมลภากรณ์ ผู้สอนที่ใช้ Python วิเคราะห์ข้อมูลและสร้างแบบจำลองพยากรณ์ในงานจริง จุดเด่นคือการสอนจากโจทย์ธุรกิจที่คุ้นเคย อธิบายหลักการเบื้องหลังแต่ละเทคนิคด้วยภาษาที่เข้าใจง่าย และพาผู้เรียนลงมือกับข้อมูลที่ไม่สมบูรณ์เหมือนข้อมูลจริง เพื่อให้กลับไปเลือกเทคนิคและวัดผลโมเดลในงานของตนเองได้อย่างมั่นใจ

ราคาค่าอบรม 9,500 บาท/ท่าน · สนใจ IN-HOUSE ขอใบเสนอราคา

9,500 บาท / ท่าน (ยังไม่รวม VAT 7%)

สมัครผ่านเว็บไซต์ ลด 10%

02-570-8449 · 088-807-9770

@itgenius · contact@itgenius.co.th

เพิ่มเพื่อน LINE