

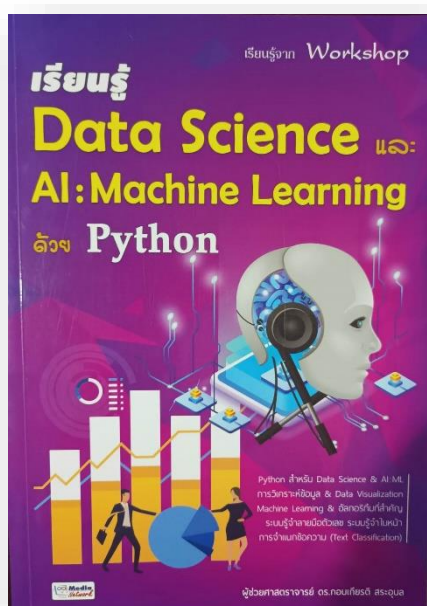
บทวิจารณ์หนังสือ: เรียนรู้ Data Science และ AI: Machine Learning ด้วย Python

ผู้เขียน: กอบเกียรติ สระอุบล

วินิตา แก้วเกื้อ

Vinita Kaewkua

สำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช



คำว่า Data Science หรือแปลเป็นภาษาไทยว่า “วิทยาศาสตร์ข้อมูล” หรือ “วิทยาการข้อมูล” ก็ตาม เป็นศาสตร์ที่ว่าด้วยการนำเอาข้อมูลมาใช้ให้เกิดประโยชน์ในด้านต่าง ๆ ไม่ว่าจะเป็น ด้านธุรกิจ

โลจิสติกส์ เศรษฐศาสตร์ การเงิน การแพทย์ วิศวกรรม และอื่น ๆ ซึ่งการนำข้อมูลมาใช้ให้เกิดประโยชน์นี้ แต่ละองค์กรสามารถเก็บรวบรวมข้อมูลได้สะดวกรวดเร็วยิ่งขึ้น ไม่ว่าจะเป็นมาจากคนที่ตอบผ่านเครื่องมือประเภทต่าง ๆ เช่น แบบสำรวจ แบบสอบถาม การสแกนแบบสอบถาม การบันทึกข้อมูลด้วยพนักงาน หรือจากระบบอัตโนมัติ เช่น จากเว็บไซต์ ระบบเซ็นเซอร์ Social Media ไฟล์เอกสาร ภาพถ่าย และอื่น ๆ ทำให้มีข้อมูลจำนวนมากที่มาจากหลายแหล่ง และมีลักษณะของข้อมูลที่แตกต่างกัน จึงเกิดเป็นแนวคิดที่ว่า จะทำอย่างไรกับข้อมูลที่มีจำนวนมากเหล่านี้เพื่อนำมาให้เกิดประโยชน์มากที่สุด ซึ่งการทำงานกับข้อมูลเหล่านี้จะมีขั้นตอนกระบวนการที่เป็นระบบเหมือนกับวิทยาศาสตร์ ดังนั้นจึงเกิดเป็นศาสตร์ที่ชื่อว่า Data Science ขึ้นมา

ส่วนคำว่า AI คำนี้หลายคนน่าจะค่อนข้างคุ้นเคยกันเป็นอย่างดี โดยคำนี้มาจากคำว่า Artificial Intelligence ที่แปลว่า ปัญญาประดิษฐ์ ซึ่งหมายถึงโปรแกรมที่ถูกสร้างและพัฒนาให้มีความฉลาด สามารถวิเคราะห์ คิดและตัดสินใจได้ โดยใช้ฐานข้อมูล ซึ่งก็คือ Data Science เป็นฐานในการตัดสินใจ ดังนั้น จึงอาจกล่าวได้ว่าทั้ง 2 คำนี้ คือ Data Science และ AI มีความเกี่ยวข้องกันอย่างหลีกเลี่ยงไม่ได้

เนื้อหาในหนังสือเล่มนี้เป็นการอธิบายที่ครอบคลุมในทั้งส่วนของ Data Science และ AI ที่กล่าวถึง Machine Learning อันเป็นสาขาย่อยแขนงหนึ่งของ AI โดยหากผู้อ่านได้อ่านทำความเข้าใจในส่วนของ Machine Learning ที่มีประเภทของการเรียนรู้ที่แตกต่างกัน 3 รูปแบบ คือ 1) Supervised Learning: การเรียนรู้แบบมีการสอน 2) Unsupervised Learning: การเรียนรู้แบบไม่ต้องการสอน และ 3) Reinforcement Learning: การเรียนรู้แบบการป้อนกลับผลลัพธ์ แล้วผู้อ่านก็จะสามารถเขียนโปรแกรมง่าย ๆ จากการไม่ใช้และใช้ Machine Learning ได้ รวมไปถึงการเขียน Model อันเป็นสมการหัวใจที่เป็นหลักสำคัญในการคิดคำนวณในแต่ละโปรแกรมของ Machine Learning ด้วย ซึ่งผู้เขียนได้เขียนวิธีการสำหรับการฝึกปฏิบัติโดยแบ่งเป็น Workshop ย่อย ๆ ตามหัวข้อต่าง ๆ ที่ผู้เขียนมีความรู้ ประสบการณ์ และความเชี่ยวชาญ จากการที่ผู้เขียนได้จบการศึกษาในระดับปริญญาเอก สาขา Information & Communication Technology for Education และปริญญาโท จากสาขา Computer Science ด้วยเกรดเฉลี่ย 4.00 และระดับปริญญาตรี ด้วยเกียรตินิยมอันดับ 1 จากสาขาวิศวกรรมเครื่องกล จากมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ (ไทย-เยอรมัน) ตลอดจนมีผลงานซอฟต์แวร์และบทความตีพิมพ์ด้านคอมพิวเตอร์ตั้งแต่ ปี พ.ศ. 2528

มาอย่างต่อเนื่อง ไม่น้อยกว่า 40 เล่ม รวมทั้งยังเป็นอาจารย์ที่มีประสบการณ์สอนกว่า 30 ปี และเป็นวิทยากรรับเชิญให้กับมหาวิทยาลัยอีกหลายแห่ง

ความน่าสนใจของหนังสือเล่มนี้ คือถ้าอ่านจากชื่อเรื่อง โดยไม่ได้พิจารณาจากเนื้อหาภายในเล่ม อาจจะคาดเดาไปได้ว่าเป็นการเขียนโปรแกรมโดยใช้ภาษา Python ซึ่งผู้อ่านหลายคนที่ไม่ได้สนใจในการเขียนโปรแกรม เพราะคิดว่าเป็นเรื่องยากและไกลตัวในการที่จะศึกษาและนำมาปฏิบัติ ก็อาจจะไม่ได้สนใจที่หยิบหนังสือเล่มนี้มาศึกษาให้เข้าใจอย่างแจ่มชัดในรายละเอียด แต่เมื่อได้เปิดดูภายในเล่มแล้ว พบว่าเป็นการปูพื้นฐานของเนื้อหาที่เกี่ยวข้องกับ Data Science และ AI อย่างครอบคลุม โดยเริ่มจาก Data Science ในรายละเอียดตั้งแต่ความหมาย ว่า Data Science หมายถึงอะไร และนำไปใช้ประโยชน์ได้อย่างไร จากการอธิบายและยกตัวอย่างด้วยภาษาที่เข้าใจง่าย เพื่อให้ผู้อ่านมองเห็นภาพไปที่ละขั้นตอน

ไม่เพียงแต่ ผู้เขียนจะกล่าวถึงรายละเอียดของ Data Science และ AI อย่างละเอียดเท่านั้น ผู้เขียนยังได้มีเนื้อหาของการติดตั้งและการเตรียมความพร้อมของคอมพิวเตอร์ในการติดตั้งโปรแกรมต่าง ๆ ที่เกี่ยวข้อง รวมถึงมีการฝึกปฏิบัติง่าย ๆ แยกตามหัวข้อย่อยเพื่อให้ผู้อ่านเข้าใจได้ง่ายจากการฝึกปฏิบัติที่ไม่ยุ่งยาก และเป็นการเสริมความรู้ความเข้าใจได้ด้วยตนเอง และมีการเพิ่มเติมเนื้อหาในส่วนของสถิติเบื้องต้นที่เกี่ยวข้องกับการใช้ Data Science และ AI เพื่อเป็นการปูพื้นฐานให้เข้าใจถึงวิธีการนำไปใช้เมื่อต้องใช้กับสถิติเบื้องต้นประเภทต่าง ๆ จากนั้นจึงเสริมความเข้าใจให้แม่นยำขึ้นด้วยการฝึกปฏิบัติอันเป็นหัวใจหลักของการเรียนรู้ในหนังสือเล่มนี้

การเรียนรู้ Data Science และ AI มีเนื้อหาที่เกี่ยวข้องอยู่หลายด้านหลายประเด็น แต่ผู้เขียนก็ได้รวบรวมประเด็นมาได้อย่างครบถ้วน โดยเริ่มจากเนื้อหาที่ง่ายไปจนถึงเนื้อหาที่ซับซ้อนขึ้น โดยในหนังสือเล่มนี้ได้นำเสนอรายละเอียดของ Data Science และ AI ให้ผู้อ่านได้ศึกษามากถึง 24 บท แต่จะยกตัวอย่างหัวข้อสำคัญ ๆ

ในแต่ละบท อันมีรายละเอียดในแต่ละบทประกอบด้วยหัวข้อ ดังนี้

บทที่ 1 บทนำ ในบทนี้จะกล่าวถึง ความหมายของ Data Science ศาสตร์และความรู้ที่จำเป็น เครื่องมือที่ใช้ในงาน Data Science งานของ Data Science ทำอะไรได้บ้าง และแนะนำการใช้หนังสือ

บทที่ 2 การติดตั้งและเตรียมความพร้อม ในบทนี้จะกล่าวถึง ชุดโปรแกรมต่าง ๆ ที่เกี่ยวกับภาษา Python คือ Anaconda และ Miniconda รู้จักกับ Python และ Jupyter Workshop 1: การใช้งาน Jupyter Lab เบื้องต้น Workshop 2: โปรแกรมตรวจสอบแพ็คเกจ และ Colab

บทที่ 3 Python สำหรับ Data Science ในบทนี้จะกล่าวถึง Python เบื้องต้น การเขียนโปรแกรมและการรัน ตัวแปร และการ Output เบื้องต้น การคำนวณพื้นฐาน ชนิดข้อมูล และอื่น ๆ

บทที่ 4 จัดการข้อมูลด้วย Pandas ในบทนี้จะกล่าวถึง รู้จักกับ Pandas DataFrame การอ่านไฟล์ข้อมูลและการบันทึก กำหนดตัวเลือกการอ่านไฟล์ Excel การตรวจสอบข้อมูลเบื้องต้น ชนิดข้อมูล และ Workshop ที่เกี่ยวข้อง

บทที่ 5 สถิติเบื้องต้น ในบทนี้จะกล่าวถึง สถิติเบื้องต้น ประชากรและกลุ่มตัวอย่าง ตัวแปร พารามิเตอร์ และสถิติประเภทของสถิติ ลักษณะของข้อมูล มาตรวัดข้อมูล การหาค่าเข้าสู่ศูนย์กลาง หรือค่ากลาง การวัดการกระจายข้อมูล การแจกแจงความถี่และ Workshop เกี่ยวข้อง

บทที่ 6 สรุปข้อมูลด้วย Pivot Table ในบทนี้จะกล่าวถึง ไฟล์ข้อมูล Excel ประกอบด้วย Workshop สรุปข้อมูลด้วย Pivot Table สร้าง Pivot Table ด้วย Pandas และ Workshop ที่เกี่ยวข้อง

บทที่ 7 Data Visualization 1 ในบทนี้จะกล่าวถึง Workshop ที่เกี่ยวข้องกับการสร้าง Data Visualization การกำหนดรายละเอียดกราฟ Seaborn ตกแต่งกราฟ กราฟ Violinplot

บทที่ 8 Data Visualization 2 ในบทนี้จะกล่าวถึง ไฟล์ข้อมูล และ Workshop ที่เกี่ยวข้อง

บทที่ 9 การจัดการข้อมูลสูญหาย ในบทนี้จะกล่าวถึง เกี่ยวกับข้อมูลสูญหาย (Missing data) ประเภทของข้อมูลสูญหาย วิธีการจัดการกับข้อมูลสูญหาย และ Workshop ที่เกี่ยวข้อง

บทที่ 10 การทำความสะอาดข้อมูล ในบทนี้จะกล่าวถึง ไฟล์ข้อมูล และ Workshop ที่เกี่ยวข้อง

บทที่ 11 ค่าผิดปกติและการกำจัด ในบทนี้จะกล่าวถึง ค่าผิดปกติ ไฟล์ข้อมูล การใช้ Z- score และ Workshop ที่เกี่ยวข้อง

บทที่ 12 ข้อมูลแบบ Time series ในบทนี้จะกล่าวถึง ข้อมูลแบบ Time series การใช้ Pandas กับ Time Series ไฟล์ข้อมูลตัวอย่าง การหาค่าเฉลี่ยและการรวม และ Workshop ที่เกี่ยวข้อง

บทที่ 13 Machine Learning เบื้องต้น ในบทนี้จะกล่าวถึง AI และ Machine Learning ประเภทของ Machine Learning, ขั้นตอน Machine Learning, Machine Learning ไม่ได้รอบรู้ไปทุกอย่าง, คำศัพท์ที่ควรรู้ และ Workshop ที่เกี่ยวข้อง

บทที่ 14 Regression ในบทนี้จะกล่าวถึง Simple Linear Regression การประเมินความแม่นยำของโมเดล, Multiple Linear Regression, Polynomial Regression, Overfitting และ Underfitting และ Workshop ที่เกี่ยวข้อง

บทที่ 15 Decision Tree ในบทนี้จะกล่าวถึง Workshop ที่เกี่ยวข้องกับ Decision Tree อย่างง่าย การติดตั้งไลบรารี แสดง Decision Tree การเข้ารหัส และ Workshop ที่เกี่ยวข้อง

บทที่ 16 การประเมิน Model ในบทนี้จะกล่าวถึง ตาราง Confusion Matrix การแบ่งข้อมูลสำหรับทดสอบ การแบ่งแบบ Cross Validation และ Workshop ที่เกี่ยวข้อง

บทที่ 17 อัลกอริทึมที่สำคัญ ในบทนี้จะกล่าวถึง Naïve Bayes Classifier, ข้อมูล Iris, Support Vector Machine, k-Nearest Neighbors, Random Forest, k-Means Clustering และ Workshop ที่เกี่ยวข้อง

บทที่ 18 ROC AUC และ Threshold ในบทนี้จะกล่าวถึง ROC curve และ AUC ค่า Threshold และ Workshop ที่เกี่ยวข้อง

บทที่ 19 ระบุรู้จำเลขลายมือเขียน ในบทนี้จะกล่าวถึง ข้อมูลภาพ และ Workshop ที่เกี่ยวข้อง

บทที่ 20 การตรวจจับและรู้จำใบหน้า ในบทนี้จะกล่าวถึง Computer Vision ตรวจจับและรู้จำใบหน้าคน OpenCV เกี่ยวกับข้อมูลภาพ และ Workshop ที่เกี่ยวข้อง

บทที่ 21 การลดมิติข้อมูล ในบทนี้จะกล่าวถึง การลดมิติข้อมูล การเลือก Feature การวิเคราะห์องค์ประกอบหลัก PCA และ Workshop ที่เกี่ยวข้อง

บทที่ 22 การปรับปรุง Model ในบทนี้จะกล่าวถึง ปรับปรุง Model เพื่ออะไร และ Workshop ที่เกี่ยวข้อง

บทที่ 23 การจำแนกหมวดหมู่ข้อความ ในบทนี้จะกล่าวถึง Text Classification การหาค่า Feature ของข้อความ และ Workshop ที่เกี่ยวข้อง

บทที่ 24 Recommender System ในบทนี้จะกล่าวถึง Recommender System และ Workshop ที่เกี่ยวข้อง

หนังสือ “เรียนรู้ Data Science และ AI: Machine Learning ด้วย Python” เป็นหนังสือที่รวบรวมข้อมูลและเนื้อหาที่เกี่ยวข้องกับ Data Science และ AI ได้อย่างโดดเด่น ด้วยเนื้อหาที่เกี่ยวข้องในทุกแง่มุมที่ผู้อ่านควรรู้ในเบื้องต้น โดยที่ไม่ได้เน้นแต่โปรแกรมทางคอมพิวเตอร์เพียงอย่างเดียว แต่ได้กล่าวถึงสถิติเบื้องต้นที่เกี่ยวข้องอันจะเป็นประโยชน์ในการเสริมความเข้าใจให้กับเนื้อหาที่เกี่ยวข้องและสามารถนำไปปฏิบัติได้จริง เป็นประโยชน์อย่างยิ่งสำหรับผู้ที่ต้องการนำความรู้ไปใช้ผ่านการฝึกฝนด้วยตนเองจาก workshop ที่ได้เรียนรู้ในแต่ละบท

เอกสารอ้างอิง

กอบเกียรติ สระอุบล. (2563). *เรียนรู้ Data Science และ AI: Machine Learning ด้วย Python*. กรุงเทพฯ: มีเดีย เนทเวิร์ค.

