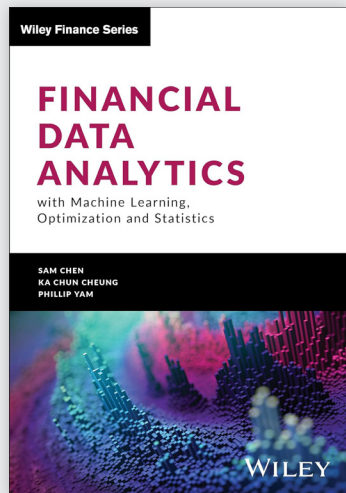




ชื่อหนังสือ : **Financial Data Analytics with Machine Learning, Optimization and Statistics**

ผู้เขียน : **Sam Chen, Ka Chun Cheung and Phillip Yam**

ปีที่พิมพ์ : **ค.ศ. 2024**

สำนักพิมพ์ : **John Wiley & Sons, Inc.**

ISBN: **978-1-119-86337-3**

การวิเคราะห์เจาะลึกข้อมูล (Data Analytics) ในโลกธุรกิจสมัยใหม่ โดยเฉพาะด้านการเงินการประกันภัยเป็นเรื่องที่ซับซ้อน เติบโตด้วยเทคโนโลยีแนวคิดวิธีการทางคณิตศาสตร์ที่ทำให้นักบัญชีที่เข้าไปทำงานเกี่ยวข้องด้วยรู้สึกสับสน หนังสือแนะนำฉบับนี้อาจเป็นตัวช่วยที่เน้นการนำเครื่องมือต่าง ๆ ไปใช้ในทางปฏิบัติพร้อมทั้งอธิบายพื้นฐานทั้งด้านคณิตศาสตร์และสถิติประกอบ

ผู้เขียนหนังสือทั้งสามท่านเป็นอาจารย์จากมหาวิทยาลัยใน Hong Kong ที่มีความเชี่ยวชาญด้านคณิตศาสตร์ สถิติ และคณิตศาสตร์ประกันภัย มีผลงานวิจัยตีพิมพ์ในวงการเป็นจำนวนมากโดยเฉพาะ Ka Chun Cheung และ Phillip Yam เมื่อมารวมตัวกันเป็นทีมผู้เขียนสหวิทยาการที่ทรงพลังทำให้หนังสือนี้เป็นเอกสารกึ่งวิชาการนำเสนอข้อมูลที่อธิบายให้เห็นขอบเขตและการประยุกต์ศาสตร์ทางสาขาการวิจัยด้านการวิเคราะห์เจาะลึกข้อมูลทางการเงินได้เป็นอย่างดี

การวิเคราะห์ข้อมูล (Data Analysis) ได้พัฒนาอย่างต่อเนื่องในแวดวงตลาดเงินตลาดทุน การประเมินความเสี่ยง เศรษฐศาสตร์มหภาค ตลอดจนการตรวจจับข้อผิดพลาดหรือความผิดปกติทางบัญชี (เช่น เทคนิค Benford's Law) การวิเคราะห์เจาะลึกข้อมูลการเงิน (Financial Data Analytics) สมัยใหม่นั้นเป็นเครื่องมือที่จำเป็นในการบรรลุผลลัพธ์ที่ละเอียดและการสร้างประโยชน์สูงสุด (Optimization) จากข้อมูลที่มีอยู่จำนวนมากผ่านทางโมเดลคณิตศาสตร์

ผู้เขียนกล่าวอย่างชัดเจนในบทนำว่าผู้อ่านหนังสือจะได้ประโยชน์เต็มที่หากมีพื้นฐานความรู้ที่แน่นในเรื่องสถิติ ความน่าจะเป็น พีชคณิตเชิงเส้น และแคลคูลัสขั้นสูง อย่างไรก็ตามเนื้อหาส่วนที่มีเครื่องหมายดอกจันแสดงถึงต้องอาศัยความเข้าใจด้านคณิตศาสตร์ที่สูงมากขึ้น ซึ่งผู้อ่านอาจข้ามไปก่อนในการอ่านรอบแรก นอกจากนั้นความคุ้นเคยกับภาษาโปรแกรมเช่น Python หรือ R ก็จะเป็นตัวช่วยในการทำความเข้าใจเนื้อหาได้อย่างดี

หนังสือเล่มนี้แบ่งเป็นสามส่วนหลัก ๆ ส่วนที่หนึ่งเกี่ยวกับ Data Cleansing and Analytical Models (มี 8 บท) ส่วนที่สองพูดถึง Linear Models (3 บท) และส่วนที่ 3 กล่าวถึง Nonlinear Models (4 บท) รวมทั้งหมด 15 บท คำโปรยที่เกริ่นช่วงต้นของบทนำโดยหยิบคำกล่าวของ Claude Shannon ที่ว่า *"We know the past but cannot control it. We control the future but cannot know it"* นับว่าน่าสนใจหากพิจารณาจากมุมมองการทำงานของนักบัญชีที่เป็นผู้เก็บรวบรวมข้อมูลธุรกรรมทางการเงินทั้งหลาย

“หนังสือเล่มนี้แบ่งเป็นสามส่วนหลัก ๆ ส่วนที่หนึ่งเกี่ยวกับ Data Cleansing and Analytical Models ส่วนที่สองพูดถึง Linear Models และส่วนที่ 3 กล่าวถึง Nonlinear Models รวมทั้งหมด 15 บท คำโปรยที่เกริ่นช่วงต้นของบทนำ โดยหยิบคำกล่าวของ Claude Shannon ที่ว่า *“We know the past but cannot control it. We control the future but cannot know it”* นับว่าน่าสนใจหากพิจารณาจากมุมมองการทำงานของนักบัญชีที่เป็นผู้เก็บรวบรวมข้อมูลธุรกรรมทางการเงินทั้งหลาย”

บทที่ 1 แนะนำเกี่ยวกับทฤษฎีการสุ่มตัวอย่างพื้นฐานและพื้นฐานของเทคนิคการจำลอง ตามด้วยบทที่ 2 ที่ทำการเปรียบเทียบระหว่างภาษา R และ Python จากนั้นบทที่ 3 และ 4 กล่าวถึงการวินิจฉัยทางสถิติสำหรับข้อมูลทางการเงิน และเครื่องมือทั่วไปทางนิติวิทยาศาสตร์ทางการเงิน (Financial Forensics) เช่น กฎของ Benford, กฎของ Zipf ที่นำมาปรับใช้เพื่อตรวจหาความผิดปกติ

บทที่ 5 เน้นการใช้การจำลอง (Simulations) เพื่อกำหนดราคาอนุพันธ์ทางการเงิน ส่วนบทที่ 6 ถึงบทที่ 8 ค่อนข้างอ่านยากเพราะเป็นหัวข้อทางเทคนิค เริ่มตั้งแต่ ก. การประมาณการทางสถิติและอัลกอริทึม Expectation-Maximization (EM) และ Majorization-Minimization (MM) ข. การคาดการณ์ความผันผวนและความสัมพันธ์แบบไดนามิกที่มีทั้งตัวแปรเดียวและหลายตัวแปร สูตรของ Kelly และ ค. การจัดการความเสี่ยงเชิงปริมาณและการสร้างแบบจำลองการพึ่งพา (Dependence Modelling) สำหรับเหตุการณ์สุดโต่ง

บทที่ 9 เป็นหัวข้อการวิเคราะห์องค์ประกอบหลัก (PCA) และระบบแนะนำพร้อมแอปพลิเคชัน ซึ่งน่าสนใจเนื่องจากนำไปประยุกต์ได้ในหลาย ๆ ด้านของธุรกิจ ส่วนบทที่ 10 อธิบายการวิเคราะห์ Regression ขั้นสูง เช่น Polynomial Regression และ Logistic Regression พร้อมอภิปรายเกี่ยวกับแบบจำลองวิธีการประเมิน เช่น Simple Receiver Operating Characteristic (ROC) Curves และพื้นที่ใต้เส้นโค้ง (Area Under Curve: AUC) สำหรับปัญหาการจำแนกประเภททั่วไป

บทที่ 11 ว่าด้วย SVM (Support Vector Machine) ซึ่งเป็นตัวจำแนกเชิงเส้น (Linear Classifier) แบบไบนารี (Binary) SVM มีประสิทธิภาพในการจำแนกข้อมูลที่มีมิติจำนวนมากได้ นอกจากนี้การใช้ฟังก์ชันเคอร์เนล (Kernel Function) ยังช่วยให้สามารถจำแนกข้อมูลที่มีความคลุมเครือได้อย่างมีประสิทธิภาพ

ส่วนที่ 3 ตั้งแต่บทที่ 12-15 เป็นโมเดล Non Linear ซึ่งการวิเคราะห์คำนวณความสัมพันธ์ หรือการตีความข้อมูลค่อนข้างที่ยากและซับซ้อน ผู้อ่านที่พื้นด้านคณิตศาสตร์หรือสถิติไม่ค่อยแข็งแรงแนะนำให้พลิกอ่านดูเฉพาะส่วนที่เป็นการประยุกต์ทางธุรกิจน่าจะเพียงพอ ตัวอย่างเช่นบทที่ 14 อาจจะอ่านเฉพาะหน้า 683-386 ที่พูดถึงการใช้ K-means Clustering ในการจำแนกข้อมูลประจำตัวลูกค้า (Customer Credential) ที่ใช้บัตรเครดิต และในการวิเคราะห์ภาพถ่ายแบ่งกลุ่มภาพถ่ายที่มีลักษณะเด่นเด่นเหมือนกัน หากเรามีร้านค้าออนไลน์ที่มีสินค้ามากมาย อาจนำ K-Means ช่วยจัดกลุ่มสินค้าให้ง่ายต่อการจัดแสดงและทำให้ผู้ซื้อเลือกสินค้าได้ง่ายขึ้น

หลังจากอ่านหนังสือเล่มนี้จบ ผู้อ่านสามารถประเมินคุณภาพข้อมูลทางการเงิน ใช้เครื่องมือวิเคราะห์ข้อมูล และนำองค์ความรู้ที่กลั่นกรองจากผลการวิเคราะห์ข้อมูลไปทำการตัดสินใจทางการเงินได้อย่างชาญฉลาด นอกจากนั้นยังเข้าใจถึงวิธีเลือกเครื่องมือวิเคราะห์ข้อมูลที่เหมาะสมสำหรับการแก้ปัญหา Regression หรือปัญหา Classification เพื่อเสริมปรับปรุงการเรียนรู้แบบ Supervised Learning ได้

มีสองอย่างที่ผมเขียนไม่ได้ครอบคลุมในหนังสือคือ Blockchains และ Deep Learning ด้วยเหตุที่ว่าเทคโนโลยีบล็อกเชนกลายเป็นประเด็นด้านความปลอดภัยทาง Internet และ Network ไปแล้วแทนที่จะเป็นเครื่องมือการวิเคราะห์ ส่วน Deep Learning จะเป็นเนื้อหาของหนังสือเล่มต่อไปของผู้เขียนทั้งสามคน

โดยสรุป หนังสือแนะนำฉบับนี้จัดเป็นคัมภีร์ (ความยาว 784 หน้า) ที่นำเสนอเนื้อหาทั้งวิธีการและการประยุกต์สำหรับนักวิจัยนักวิชาการ ผู้ปฏิบัติงานด้านการวิเคราะห์เจาะลึกข้อมูล นักศึกษาปริญญาโทและเอก สามารถใช้เป็นกรอบการทำงานอย่างเป็นระบบสำหรับการขยายผลในอนาคตเพื่อเพิ่มมูลค่าการวิเคราะห์ข้อมูลทางการเงิน

