

ฉบับที่ 24,921 วันศุกร์ที่ 5 มกราคม พ.ศ. 2561 หน้า 23



หนึ่ง-ศูนย์-ศูนย์-หนึ่ง

ดีไอวายวิทยาการข้อมูล

สวัสดิ์ปีใหม่ 2561 ครับท่านผู้อ่านทุกท่าน

จาก 1001 เมื่อหลายปีที่แล้ว ที่ผมเกริ่นถึงเทคโนโลยีทางด้านปัญญาประดิษฐ์ (หรือเอไอ) ที่กำลังจะเกิดขึ้นในปีี้ เพื่อให้เป็นแนวทางสำหรับองค์กรต่าง ๆ ในประเทศได้ปรับตัวเตรียมนำเทคโนโลยีเหล่านี้มาใช้

ผมจึงวางแผนไว้ว่าในปีี้ คอลัมน์ 1001 จะพาทุกท่านที่สนใจ หันมาลงมือใช้ ลงมือทำเพื่อให้คอลัมน์ 1001 กลายเป็นจุดเริ่มต้นที่รวดเร็ว สำหรับน้อง ๆ นักเรียน นิสิต นักศึกษา ผู้ปฏิบัติงานในองค์กรต่าง ๆ จะได้รับเวลาในการศึกษาเครื่องมือขั้นต้น

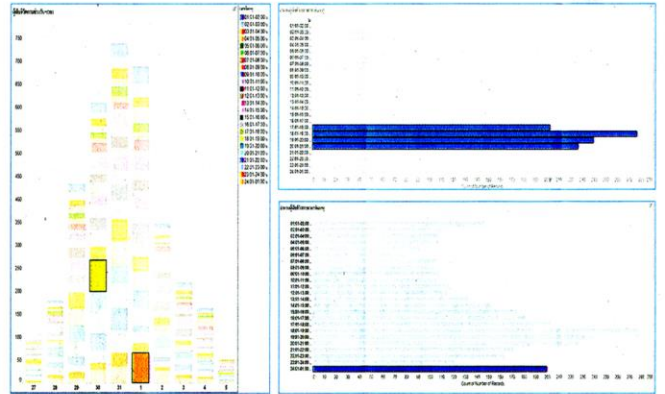
เราประเดิมกันด้วยดีไอวาย (DIY หรือ Do It Yourself) ลงมือทำมันด้วยตัวเองกับวิทยาการข้อมูลกันดีกว่าครับ

ผมมีเครื่องมือที่ใช้เวลาไปบรรยายให้กับหน่วยงานต่าง ๆ อยู่สามชิ้นที่ทุกท่านสามารถทดลองใช้ หาข้อมูลเพื่อเรียนรู้ได้ด้วยตัวเอง และที่สำคัญคือ ไม่ต้องตั้งงบประมาณหลายสิบล้านมาจัดซื้อให้เปลือง (เพราะเครื่องมือเหล่านี้ ฟรี หรือเสียค่าใช้จ่ายในกรณีที่ต้องการคุณลักษณะพิเศษเพิ่มขึ้น)

เครื่องมือทั้งสามชิ้นที่ว่านี้ คือ ภาษาโปรแกรมอย่างไพทอน (python) รัชภัษผู้รักสำหรับนักเขียนโปรแกรมทุกระดับ เครื่องมือแสดงข้อมูล (Data Visualization) อย่างแท็บโบล (Tableau) ที่ปัจจุบัน อนุญาตให้ใช้ฟรีในรุ่นแท็บโบลสาธารณะ (Tableau Public) และเครื่องมือสำหรับใช้เพื่อหารูปแบบที่ซ่อนอยู่ในข้อมูลอย่าง แรพพิคไมเนอร์ (RapidMiner) ที่มีเครื่องมือเพื่อการวิเคราะห์ข้อมูลที่ลึกซึ้งกว่า (แน่นอนว่า ต้องเรียนรู้กันเล็กน้อยจึงจะสามารถใช้มันได้คล่อง)

สำหรับทุกท่านที่เป็นเจ้าของข้อมูลอยู่แล้ว ท่านสามารถใช้เครื่องมือที่ท่านคุ้นเคย ไม่ว่าจะเป็นเอ็กเซล แท็บโบล ไพทอน หรือแรพพิคไมเนอร์ ในการวิเคราะห์ข้อมูลได้ง่าย ๆ เพื่อ “สนับสนุน” การตัดสินใจตามความจริงจากข้อมูล และผมค่อนข้างมั่นใจว่า ในปัจจุบันการออกนโยบายต่าง ๆ จากภาครัฐ ทำอยู่บนการ “สนับสนุน” จากข้อมูล

วันนี้ผมหยิบตัวอย่างจากเครื่องมือแท็บโบลสาธารณะมาให้ท่านผู้อ่าน (โดยเฉพาะอย่างยิ่ง องค์กรเจ้าของข้อมูลและผู้เกี่ยวข้องกับข้อมูลนั้น ๆ ได้หันมาสนใจ และทดลองนำมาใช้กับองค์กรของท่าน)



รูปภาพทั้งสามรูปนี้ (เป็นรูปเก่าที่ผมเคยเล่าถึงรายละเอียดการวิเคราะห์ข้อมูลชุดนี้ไปในแทบทุกช่วงเวลา 7 วัน อันตราย ผ่านคอลัมน์ 1001 ไปแล้วครับ) เป็นตัวอย่างที่ท่านสามารถนำเข้าไปไฟล์เอ็กเซลและแสดงผลได้ในเวลาไม่เกิน 10 นาที ถ้าท่านต้องการตอบคำถามว่า ช่วงเวลาใดคือช่วงเวลาอันตรายที่สุดในการขับขีช่วง 7 วันอันตรายของปีใหม่ (ปี 2551-2558 ข้อมูลจาก data.go.th) คำตอบคือช่วงเวลาโพลีเพล็กซ์วันที่ 30 ธันวาคม และช่วงเวลาหลังการนับถอยหลัง (หลังเที่ยงคืนถึงตีหนึ่ง)

ผมดีใจที่เห็นนโยบายภาครัฐใหม่ ๆ ออกมาทุกปี เพื่อพยายามลดจำนวนผู้บาดเจ็บและเสียชีวิตบนท้องถนน และหลายครั้ง นโยบายเหล่านั้นสอดคล้องกับข้อมูลที่มีอยู่จริง ๆ (ไม่ให้นั่งท้ายกระบะ เพราะบนทางหลวงผู้เสียชีวิตที่เป็นผู้โดยสารรถกระบะมีจำนวนสูงสุด) และบางครั้งมีการออกมาบอกว่าจะเก็บข้อมูลให้มากขึ้น เพราะข้อมูลที่มีอยู่ยังไม่เพียงพอ เช่น ความเร็วรถขณะเกิดเหตุ (หลังจากวิเคราะห์ทางนิติวิทยาศาสตร์แล้ว) ตำแหน่งของอุบัติเหตุ (จะรู้ได้ว่าถนนตรงไหนมีปัญหาอะไร ถึงทำให้เกิดอุบัติเหตุจนได้รับบาดเจ็บและเสียชีวิต)

ขอให้ปีนี้เป็นปีที่ผู้ปฏิบัติงานในองค์กรเริ่มหันมาใช้เครื่องมือเพื่อการวิเคราะห์ข้อมูลของตัวเองได้เชี่ยวชาญขึ้น ตัดสินใจได้อย่างถูกต้องและรวดเร็วมากขึ้นนะครับ.

สุกรี สันสุภัญญ์

(sukree.s@chula.ac.th)

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย