

แบบฝึกหัด Machine Learning

โดย อ.ดร.เสกฐวิทย์ เกิดผล

โจทย์

ให้นักลิตนำเข้าไฟล์ชื่อ census-income-train.csv แล้วทำการสร้าง โมเดลเพื่อทำนายว่าผู้เสียภาษีคนไหนมีรายได้อ่ปีมากกว่า \$50,000 ให้ใช้กระบวนการในการสร้างและทดสอบโมเดลที่ถูกต้อง ลองใช้หลาย algorithm และหาวิธีปรับจนให้โมเดลมีความแม่นยำมากขึ้น ระวังเรื่องการ overfit ด้วย

เมื่อนิสิตส่งไฟล์ notebook มาแล้ว อาจารย์จะนำข้อมูลทีกันไว้มาทดสอบกับ โมเดลของนิสิต การให้คะแนนจะดูจากกระบวนการเท่านั้น โดยไม่ดูจากความแม่นยำของ โมเดล

คำอธิบายข้อมูล

age: อายุของผู้เสียภาษี

workclass: ประเภทอาชีพ

weight: ไม่มีคำอธิบาย

education: ระดับการศึกษา

edu num: จำนวนปีที่ใช้ในการศึกษา

marital status: สถานะการแต่งงาน

occupation: อาชีพ

relationship: ความสัมพันธ์กับผู้อาศัยร่วม

race: เผ่าพันธุ์

sex: เพศ

capital-gain: จำนวนเงินที่ได้จากการลงทุน

capital-loss: จำนวนเงินที่เสียไปจากการลงทุน

hours-per-week: จำนวนชั่วโมงต่อสัปดาห์ที่ใช้ทำงาน

native country: ประเทศที่เกิด

label: Yes คือมีรายได้อ่มากกว่า \$50,000 No คือได้น้อยหรือเท่ากับ \$50,000