

LAB 2 Data Preprocessing

040613702 Machine Learning

จากชุดข้อมูลในไฟล์ pima_diabelete.csv

1. จงเขียนโปรแกรมเพื่ออ่านข้อมูล pima_diabelete.csv และแสดงผลลัพธ์ 20 เรคอร์ดแรก
2. จงเขียนโปรแกรมเพื่อแสดงข้อมูลผู้ป่วยที่มีอายุมากกว่า 30 และมีความหนาของผิวหนัง (skin) มากกว่า 35 มิลลิเมตรขึ้นไป
3. ลบคอลัมน์ Unnamed:0 ออกจากข้อมูลพร้อมแสดงผลลัพธ์
4. กำจัด missing values โดยใช้วิธี replace by means
5. จงปรับค่า feature 'plas' และ 'skin' ให้มีค่าอยู่ระหว่าง 0 ถึง 1
6. จงปรับค่า feature 'insu' ให้อยู่ในรูปของค่าคะแนน Z
7. แสดงเมตริกซ์ correlation ของชุดข้อมูล
8. แสดงรายการ feature ที่มีค่า correlation กับแอตทริบิวต์ 'class' มากที่สุด 3 อันดับแรก
9. จงสร้างชุดข้อมูลใหม่ โดยใช้ผลลัพธ์ของรายการ features ที่ได้จากในข้อที่ 8 และแอตทริบิวต์ class พร้อมบันทึกในรูปแบบไฟล์ csv ตั้งชื่อ new_data.csv
10. บันทึกไฟล์ Lab2_รหัสนักศึกษา.ipynb อัปโหลดใน Google Classroom