

รายงาน  
เรื่อง Mini Project - Customer Classification

เสนอ

ผศ.ดร. คณบดี ศรีสมบูรณ์

จัดทำโดย

นาย ณัฐนันท์ วารีวณิชพันธ์ 6201011631056  
นาย นาวิน โพธิ์เย็น 6201011631099

รายงานนี้เป็นส่วนหนึ่งของรายวิชา Machine Learning ( 010113713 )

ภาคเรียนที่ 1 ปีการศึกษา 2565

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

Table

Description automatically generatedDataSet ที่เลือกในการทำ

Table

Description automatically generated

Table

Description automatically generated**Data Set**  
ข้อมูลที่ได้มาจากงานต้นแบบ ไฟล์ CustomerSegmentationClassification.csv

Table

Description automatically generated  
ให้ Data frame อ่านค่าจาก url ของไฟล์ CustomerSegmentationClassification.csv แล้วแสดงข้อมูล

Graphical user interface, application

Description automatically generatedเช็คตรวจสอบข้อมูล

Table

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface

Description automatically generated with medium confidenceจากการตรวจสอบข้อมูลจะมีบางค่าที่หายไปจาก DataSet

A picture containing table

Description automatically generatedทำการแก้ไข Ever\_Married โดยการเติมคำลงไปด้วยการ Random [Yes , No]

Graphical user interface, application

Description automatically generatedตรวจสอบค่าเป็นไปตามที่ต้องการ

Table

Description automatically generated with medium confidenceทำการแก้ไข Graduated

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, application

Description automatically generatedทำการแก้ Profession โดยทำการนับประเภทก่อนว่ามีอะไรบ้าง

Table

Description automatically generated with low confidence

Graphical user interface, application, Word

Description automatically generated

ทำการแก้ Work\_Experience โดยทำการนับประเภทก่อนว่ามีอะไรบ้าง

Graphical user interface, application

Description automatically generated

Table

Description automatically generated with low confidence

Graphical user interface, application, Word

Description automatically generated

ทำการแก้ Family\_Size โดยทำการนับประเภทก่อนว่ามีอะไรบ้าง

Graphical user interface

Description automatically generated with medium confidence

A picture containing table

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

ทำการแก้ Var\_1 โดยทำการนับประเภทก่อนว่ามีอะไรบ้างGraphical user interface, application

Description automatically generated

A picture containing graphical user interface

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, application, Word

Description automatically generatedเมื่อเติมจนครบแล้ว ทำการตรวจสอบประเภทข้อมูลก่อนว่ามีประเภทอะไรบ้าง เพราะส่วนใหญ่ในการทำงานควรจะเป็นประเภท int64

Graphical user interface, application, Word

Description automatically generatedTable

Description automatically generated with medium confidenceเลือกที่ทำการแปลงง่ายก่อนคือ จาก float64 เป็น int64

แต่การจะแปลงจาก object ให้เป็น int64 ไม่ได้จึงทำการกำหนดค่าให้แต่ละคำใน column แทนเลขไปแทน เช่น A=1 , B=2 , C=3 เป็นต้นเพื่อให้สะดวกในการใช้งาน  
Column ที่จะทำการแปลง มีดังนี้  
 Gender , Ever\_Married , Graduate โดย Yes = 1 , No = 0  
 Spending\_Score โดย High = 2 , Average = 1 , Low = 0  
 Column ที่เหลือก็แทนไล่ไปจาก 0

Table

Description automatically generated

A picture containing table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated with low confidence

Table

Description automatically generated

Table

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

A picture containing table

Description automatically generated

  
ทำการ Export file เพื่อนำไปใช้งานเป็น CSV

จะทำการ Normalization ค่าในตารางในไฟล์ล่าสุด แต่ค่าใน Column Segmentation และ ID ไม่จำเป็นจึงจะทำการตัดทิ้งไว้ก่อน

Table

Description automatically generated with medium confidence

ทำการดูค่า Mean , Max , Min เทียบจาก Segmentation

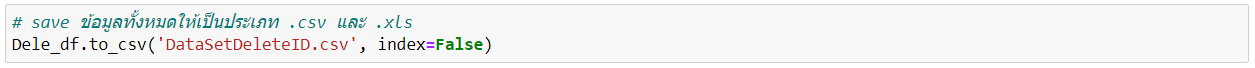
Table

Description automatically generatedA picture containing diagram

Description automatically generated

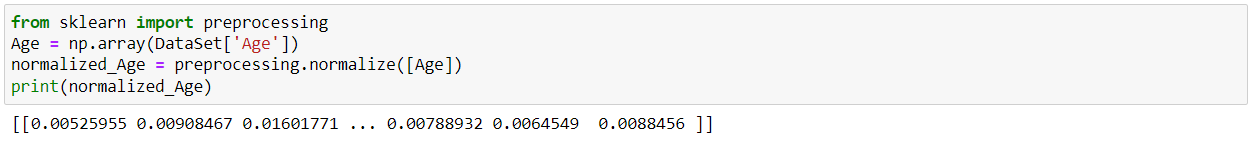
Table

Description automatically generated

ทำการ Export file เพื่อนำไปใช้งานเป็น CSV

Table

Description automatically generatedเปรียบเทียบค่าจากการที่ตัด กับไม่ตัดเพื่อเตรียมนำไปทำ Normalization

การ Normalization จากสูตรใน Column เดียว

ทำการ Normalization ทั้ง DataSet แล้วเปรียบเทียบทั้งค่าเก่าและค่าใหม่

Table

Description automatically generated A picture containing graphical user interface

Description automatically generated

Table

Description automatically generatedดูภาพรวมของ Dataset แล้วทำการ Export



Graphical user interface, application, Word

Description automatically generatedแต่ตัว DataSet ยังไม่สมบูรณ์ยังขาดตัว Segmentation ที่ทำการตัดไปและ ID แต่ตัว ID ผมคิดว่าไม่จำเป็นจึงไม่ทำการนำมาใส่คืน ก่อนอื่นก็ทำการแยก file ให้เหลือแค่ Segmentation แล้วนำมารวม  
 Table

Description automatically generated



Table

Description automatically generatedตรวจสอบ file ที่จะนำมารวมกัน

Table

Description automatically generatedเท่านี้ก็ได้ File ที่พร้อมในการนำไป Train ข้อมูลให้กับ Model เรียบร้อย

**KNN**

กลุ่มของพวกเราเลือกที่จะใช้วิธี KNN เป็นกระบวนการ Learning Algorithm แรก เนื่องจาก DataSet ที่กลุ่มพวกเราเลือกใช้ เป็นการจัดกลุ่ม CustomerSegmentation ด้วยการแบ่งและ  
เป็น Machine Learning ประเภท Supervised แบบ Classification นั้นคือ จำเป็นจะต้องมี DataSet ที่มีเฉลย (Label) ให้ด้วย ก็จะตรงเงื่อนไขทุกอย่าง อีกทั้งยังง่ายและสะดวกต่อการทำและทดลองเพราะมีความง่าย ในระดับที่เข้าใจได้และสมการก็เข้าใจไม่ยาก เพียงแต่ตัวข้อมูลควรจะ Normalization ก่อนจึงจะได้ประสิทธิภาพที่ดีขึ้น



A picture containing table

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Diagram, schematic

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated ใช้การดู Accuracy ของแต่ละ Feature ก่อนแล้วจึงตัดสินใจเลือก โดยจะควบคุมให้ตัวแปรมีลักษณะใกล้เคียงกันทุกอย่าง เพื่อจะนำมาเปรียบเทียบกัน

Graphical user interface, application

Description automatically generatedใช้การจัดกลุ่มข้อมูล Classification ด้วย KNN โดยใช้ library ของ Sklearn

Graphical user interface, application

Description automatically generatedแบ่ง x\_train และ y\_train เพื่อนำไปใช้กับ Model

สร้างแบบจำลอง K-Nearest Neighbours ด้วยคลาส KNeighborsClassifier ใน Sklearn

Graphical user interface, application

Description automatically generatedFeature : Gender

Background pattern

Description automatically generated

Graphical user interface, application

Description automatically generatedFeature : Ever\_Married

Background pattern

Description automatically generated

Graphical user interface, application

Description automatically generatedFeature : Age

Background pattern

Description automatically generated

Graphical user interface, application, Word

Description automatically generatedFeature : Graduated

Background pattern

Description automatically generated with low confidence

Graphical user interface, application, Word

Description automatically generatedFeature : Profession

A picture containing background pattern

Description automatically generated

Graphical user interface, application, Word

Description automatically generatedFeature : Work\_Experience

Background pattern

Description automatically generated

Graphical user interface, application, Word

Description automatically generatedFeature : Spending\_Score

Background pattern

Description automatically generated

Graphical user interface, application

Description automatically generatedFeature : Family\_Size

Background pattern

Description automatically generated

Graphical user interface, application

Description automatically generatedFeature : Var\_1

A picture containing background pattern

Description automatically generated

Graphical user interface

Description automatically generated with medium confidenceจากการทดสอบทีละ Feature จะได้ค่า Accuracy ที่แตกต่างกัน จะใช้การเลือกค่าตัวหลักในการนำมาใช้ Train ข้อมูลคือ Feature : Age และทำการหา Feature มาเพิ่มเนื่องจากค่า Accuracy ยังไม่เป็นที่น่าพอใจ ใช้หลักการเดิมในการทำ

Graphical user interface

Description automatically generated with medium confidenceFeature : Age , Gender

Background pattern

Description automatically generated

Graphical user interface, application, Word

Description automatically generatedFeature : Age , Ever\_Married

Background pattern

Description automatically generated

Graphical user interface, application

Description automatically generatedFeature : Age , Graduated

A picture containing background pattern

Description automatically generated

Graphical user interface, application

Description automatically generatedFeature : Age , Profession

Background pattern

Description automatically generated

Graphical user interface, application, Word

Description automatically generatedFeature : Age , Work\_Experience

Background pattern

Description automatically generated

Graphical user interface

Description automatically generated with medium confidenceFeature : Age , Spending\_Score

Background pattern

Description automatically generated

Graphical user interface, application, Word

Description automatically generatedFeature : Age , Family\_Size

Background pattern

Description automatically generated with medium confidence

Graphical user interface, application, Word

Description automatically generatedFeature : Age , Var\_1

Background pattern

Description automatically generated with medium confidence

จากการทดสอบนำ Feature มาใช้ร่วมกัน 2 Feature จะได้ค่า Accuracy ที่เพิ่มขึ้นแต่ก็จะสังเกตุเห็นได้ว่าบาง Feature เมื่อทำการทดสอบเพียงตัวเดียวก็จะทำให้ได้ค่า Accuracy ที่น้อยแต่เมื่อนำมารวมกันกับ Feature อื่นๆค่า Accuracy กลับมากกว่าบาง Feature ซึ่งก็เนื่องมาจากความสัมพันธ์ของ Feature แต่ละตัวนั้นแตกต่างกัน อีกปัจจัยหนึ่งคือมี Feature ที่ได้ค่า Accuracy ที่มีค่ามากในการ Train แต่จะมีค่าที่ลดลงในการ Test โดย Feature ลักษณะนั้นจะไม่นำมาในการทดสอบเพราะอาจเกิดปัญหา Overfitting ได้ในอนาคต โดย Feature ที่จะใช้ร่วมกันคือ Age , Spending\_Score จากนั้นทำการหา Feature มาเพิ่มเนื่องจากค่า Accuracy ยังไม่เป็นที่น่าพอใจ ใช้หลักการเดิมในการทำ

Graphical user interface, application

Description automatically generated3 Feature

Graphical user interface, application

Description automatically generated4 Feature

A picture containing text

Description automatically generated  
5 Feature

Graphical user interface, application

Description automatically generated Feature : Age , Spending\_Score , Gender

Background pattern

Description automatically generated

Graphical user interface, application, Word

Description automatically generated Feature : Age , Spending\_Score , Ever\_Married

Background pattern

Description automatically generated

Graphical user interface, application

Description automatically generated Feature : Age , Spending\_Score , Graduated

Background pattern

Description automatically generated

Graphical user interface, application

Description automatically generatedFeature : Age , Spending\_Score , Profession

Background pattern

Description automatically generated

Graphical user interface, application

Description automatically generatedFeature : Age , Spending\_Score , Work\_Experience

Background pattern

Description automatically generated

Graphical user interface, application

Description automatically generatedFeature : Age , Spending\_Score , Family\_Size

Background pattern

Description automatically generated

Graphical user interface, application, Word

Description automatically generatedFeature : Age , Spending\_Score , Var\_1

A picture containing background pattern

Description automatically generated

ในกรณีที่ได้ค่า Accuracy ที่ดีที่สุดคือ Feature : Age , Spending\_Score , Profession

Graphical user interface, application

Description automatically generated**ทดสอบ กรณี 4 Feature**  
Feature : Age , Spending\_Score , Profession , Gender

Background pattern

Description automatically generated

Graphical user interface

Description automatically generatedFeature : Age , Spending\_Score , Profession , Ever\_Married

Background pattern

Description automatically generated with medium confidence

Graphical user interface, text, application

Description automatically generatedFeature : Age , Spending\_Score , Profession , Graduated

A picture containing background pattern

Description automatically generated

Chart, scatter chart

Description automatically generatedFeature : Age , Spending\_Score , Profession , Work\_Experience

Background pattern

Description automatically generated

Graphical user interface, application

Description automatically generatedFeature : Age , Spending\_Score , Profession , Family\_Size

Background pattern

Description automatically generated

Graphical user interface, application, Teams

Description automatically generatedFeature : Age , Spending\_Score , Graduated , Var\_1

Background pattern

Description automatically generated with medium confidence

หลังจากทำการทดสอบเพิ่มเกิน 4 Feature เข้าไปได้ค่า Accuracy ที่มากกว่า Feature 4 ตัวที่ประกอบด้วย Feature : Age , Spending\_Score , Profession , Gender อยู่เล็กน้อยจึงพิจารณาใช้ Feature 5 ตัวนี้เพื่อเตรียมตัวไปปรับค่า Parameter ต่างๆ

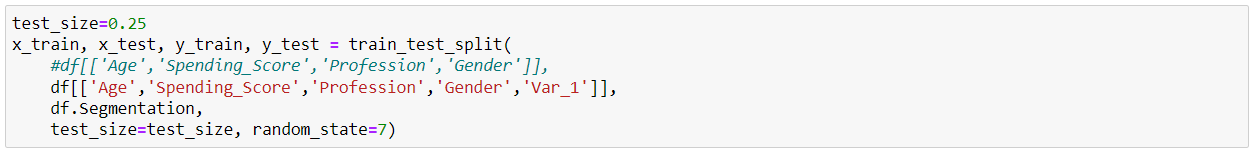
Text, letter

Description automatically generatedText

Description automatically generated

5 Feature

4 Feature

ปรับเปลี่ยน test\_size เพื่อหาช่วงที่ ค่า Accuracy เพิ่มขึ้นแต่ไม่มีช่วงที่เพิ่มเลยจึงใช้ คือ 0.25

Background pattern

Description automatically generated

Graphical user interface, text, application, email

Description automatically generatedปรับเปลี่ยนค่า Prefer\_k เพื่อหาช่วงที่ ค่า Accuracy เพิ่มขึ้น   
แต่เมื่อทำการลด ค่า K ลงเป็น 16 จะได้ค่า Accuracy ที่เพิ่มขึ้นเล็กน้อย จึงใช้ค่า K = 16

Text

Description automatically generatedทำการทดลองทดสอบประสิทธิภาพแบบจำลองด้วย cross validation

Graphical user interface, text, application, email

Description automatically generatedแสดงค่าประสิทธิภาพของ Model

Graphical user interface, text, application

Description automatically generatedทดสอบการจำลองการทำงานของ Model โดยการป้อนข้อมูลเพื่อให้ model คาดเดาโดยจะใส่ทั้งข้อมูลจริงและข้อมูลที่ปรับเปลี่ยนค่านิดหน่อย

Graphical user interface, text, application, email

Description automatically generated

โดยข้อมูลในชุดที่ 0 , 2 นั้นเป็นข้อมูลจริงจากตารางและให้ผลเฉลยตรง ส่วนข้อมูลชุดที่ 1 ใส่ข้อมูลที่สร้างเองและสร้างผลเฉลยเอง ส่วนข้อมูลชุดที่ 3 เป็นการตัดข้อมูลบางส่วนทิ้งจากข้อมูลจริงและให้ผลเฉลยเป็นจริง ส่วนชุดสุดท้ายชุดที่ 4 เป็นข้อมูลจริงจากตารางและให้ผลเฉลยผิด

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application

Description automatically generatedทำการทดลองให้ Model ทำการหาค่า k ที่เหมาะสมที่สุดเอง

Graphical user interface

Description automatically generatedกำหนดช่วงเริ่มตั้งแต่ 1 ถึง 50 จากนั้นสร้างตารางเพื่อดูรายละเอียดต่างๆ

Graphical user interface, application

Description automatically generated

Graphical user interface

Description automatically generated with medium confidenceจากการสร้าง กราฟแสดงจำนวนของค่า K เทียบกับค่า Accuracy นั้นตั้งแต่ช่วง 1 ถึง 50 จะสามารถทราบช่วงที่ได้ค่า Accuracy ที่สูงที่สุดนั้นคือค่า K ที่ 24

Graphical user interface, text, application, email

Description automatically generatedเปรียบกับค่า Accuracy ของ Model ที่ได้กำหนดขึ้นมาเองโดยมีค่า K = 16

**ANN** กลุ่มของพวกเราเลือกที่จะใช้วิธี KNN เป็นกระบวนการ Learning Algorithm แรก เนื่องจาก DataSet ที่กลุ่มพวกเราเลือกใช้ เป็นการจัดกลุ่ม CustomerSegmentation ด้วยการแบ่งและ  
เป็น Machine Learning ประเภท Supervised แบบ Classification นั้นคือ จำเป็นจะต้องมี DataSet ที่มีเฉลย (Label) ให้ด้วย ก็จะตรงเงื่อนไขทุกอย่าง อีกทั้งยังง่ายและสะดวกต่อการทำและทดลองเพราะมีความง่าย ในระดับที่เข้าใจได้และสมการก็เข้าใจไม่ยาก เพียงแต่ตัวข้อมูลควรจะ Normalization ก่อนจึงจะได้ประสิทธิภาพที่ดีขึ้น