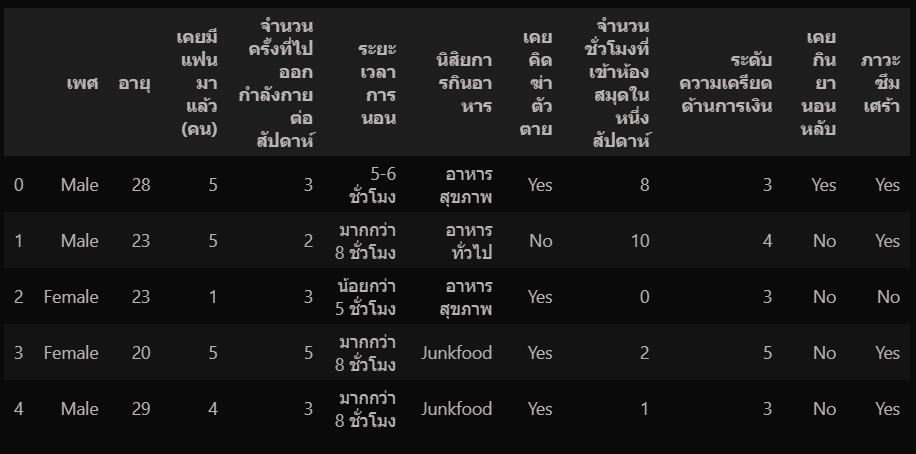
**การนำเข้าข้อมูลและการเตรียมข้อมูล**

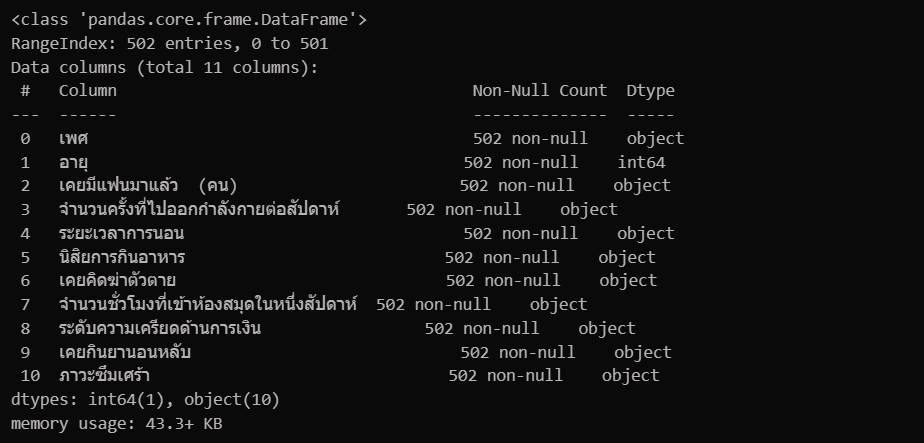
* โดยในชุดข้อมูลประกอบไปด้วย feature ดังนี้

|  |  |
| --- | --- |
| **features** | **Description** |
| เพศ | ระบุเพศ |
| อายุ | ระบุอายุ |
| เคยมีแฟนมาแล้ว (คน) | จำนวนแฟนที่เคยมีมาแล้ว |
| จำนวนครั้งที่ไปออกกำลังกายต่อสัปดาห์ | จำนวนครั้งความถี่ในการออกกาลังกายในหนึ่งสัปดาห์ |
| ระยะเวลาการนอน | ช่วงเวลาการนอนหลับ |
| นิสัยการกินอาหาร | ประเภทอาหารที่ชอบรับประทาน |
| เคยคิดฆ่าตัวตาย | ระบุว่าเคยมีความคิดฆ่าตัวตายหรือไม่ |
| จำนวนชั่วโมงที่เข้าห้องสมุดในหนึ่งสัปดาห์ | ระบุเวลาที่ใช้ในห้องสมุดต่อสัปดาห์ |
| ระดับความเครียดด้านการเงิน | ระดับความเครียดเกี่ยวกับการเงิน |
| เคยกินยานอนหลับ | ระบุว่าเคยใช้ยานอนหลับหรือไม่ |
| ภาวะซึมเศร้า | สถานะภาวะซึมเศร้า |

* แสดงชุดข้อมูลที่นำเข้ามา



* การแสดงรายละเอียดของชุดข้อมูล



* แสดงรายละเอียดของ features target

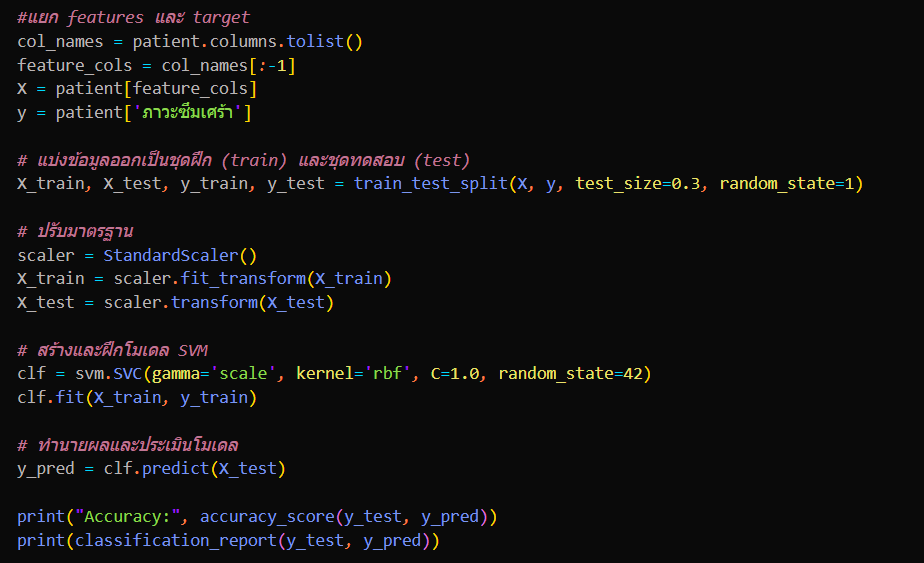


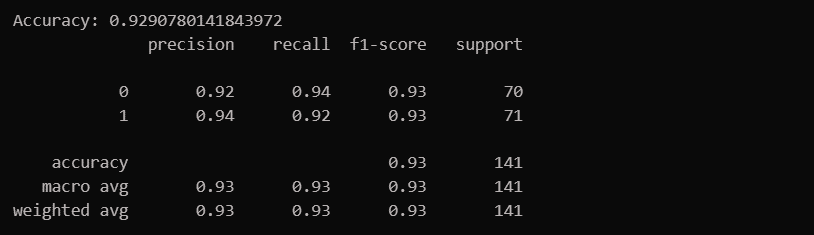
* แปลงข้อมูลที่ไม่เป็นตัวเลขให้เป็นค่าที่เหมาะสม โดย Label Encoding และตรวจสอบข้อมูลที่หายไปหรือไม่สมบูรณ์แล้วแสดงชุดข้อมูลอีกครั้งหลังจากทำการแปลงข้อมูล



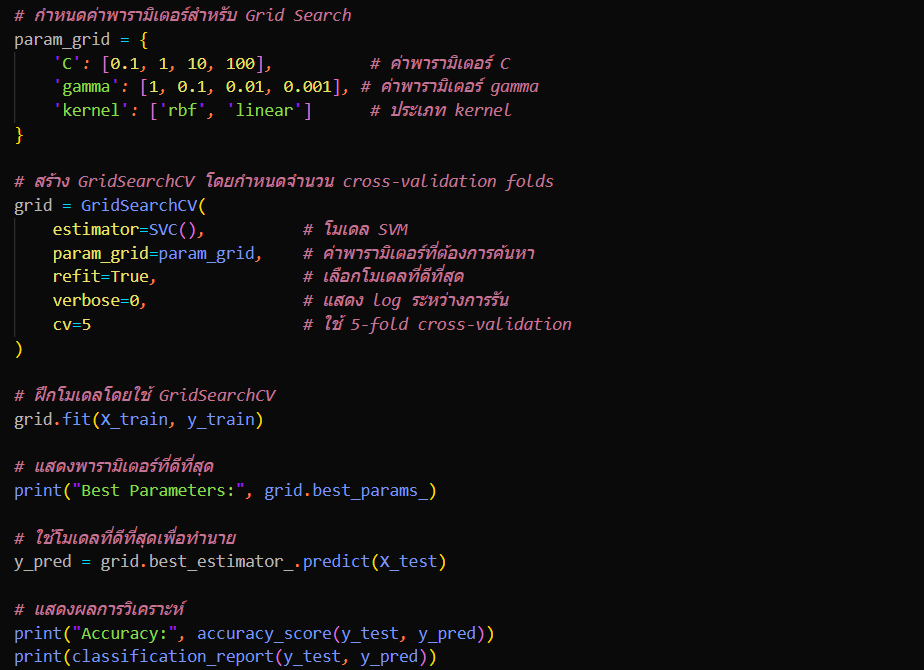
**การทำนายผล**

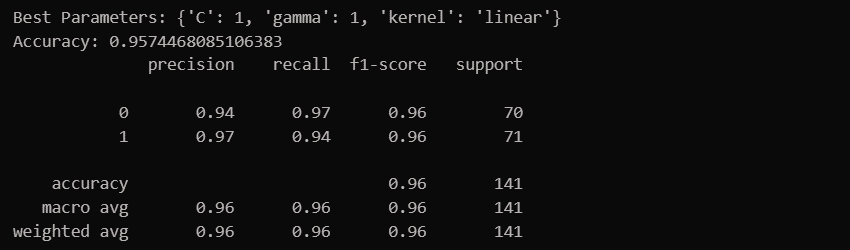
* การทำนายครั้งแรก



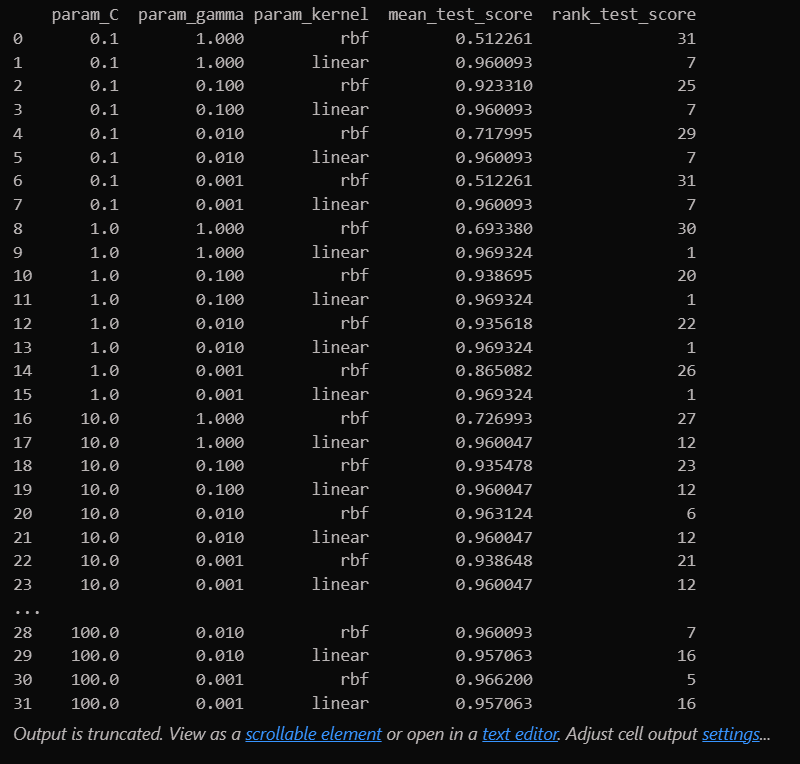
****

* ลองใช้ GridSearchCV โดยจะค้นหาค่าที่ดีที่สุดสำหรับพารามิเตอร์โดยการลองค่าที่กำหนดในแต่ละตัวเลือก

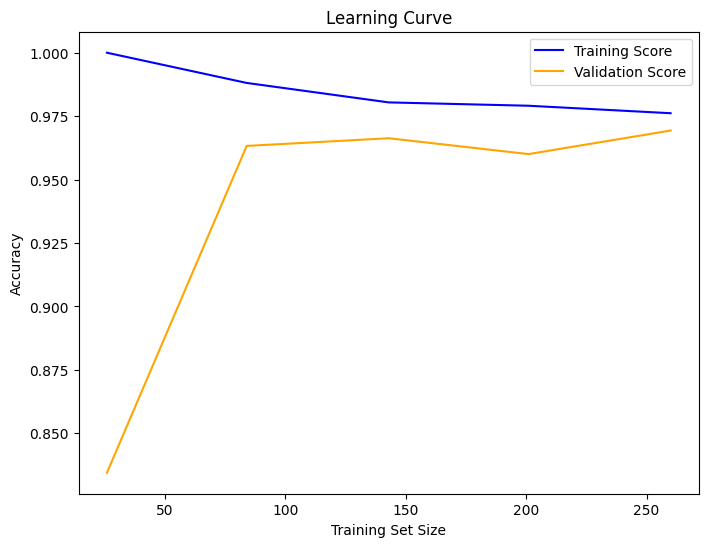




และมีการแสดงผลลัพธ์ทั้งหมดที่ได้จากการค้นหา



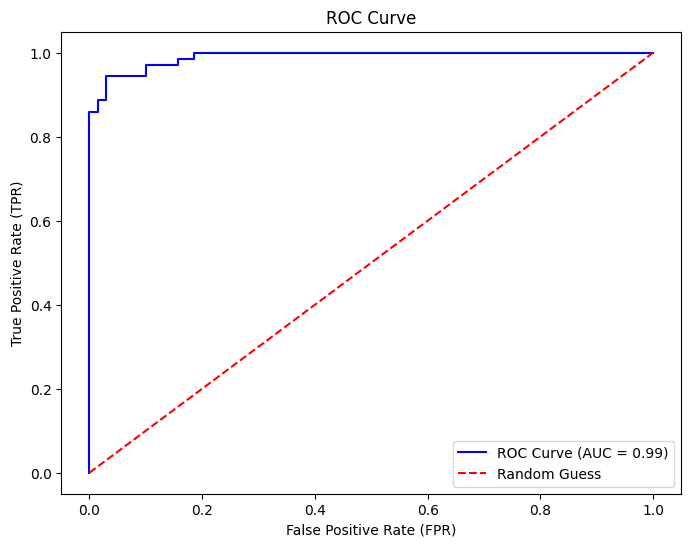
* แสดง Learning Curve



1. **แกน X** แสดงจำนวนข้อมูลที่ใช้ในการฝึก (Training Set) ซึ่งเพิ่มขึ้นเรื่อย ๆ จากซ้ายไปขวา
   * เริ่มจากใช้ข้อมูลจำนวนน้อย (เช่น 50 ตัวอย่าง) ไปจนถึงข้อมูลทั้งหมดในชุดฝึก (เช่น 250 ตัวอย่าง)
2. **แกน Y** แสดงค่าความแม่นยำ (Accuracy) ของโมเดล ซึ่งมีค่าระหว่าง 0 ถึง 1.0
   * ค่าใกล้ 1.0 หมายถึงโมเดลมีความแม่นยำสูง
   * ค่าใกล้ 0 หมายถึงโมเดลมีความแม่นยำต่ำ
3. **เส้นสีน้ำเงิน (Training Score)** แสดงค่าความแม่นยำของโมเดลบนชุดข้อมูลฝึก (Training Set)
   * ในช่วงเริ่มมีค่าใกล้ 1.0 เพราะโมเดลสามารถจดจำข้อมูลฝึกทั้งหมดได้แต่เริ่มลดลงเล็กน้อย เพราะข้อมูลฝึกเพิ่มขึ้น ทำให้โมเดลต้องปรับตัวให้เหมาะสมมากขึ้น ซึ่งยังคงอยู่ในระดับสูง (>0.95) หมายความว่าโมเดลสามารถเรียนรู้ข้อมูลฝึกได้ดี
4. **เส้นสีส้ม (Validation Score)** แสดงค่าความแม่นยำของโมเดลบนชุดข้อมูล (Validation Set)
   * ค่าอยู่ในระดับต่ำ (<0.9) เพราะโมเดลไม่สามารถ generalize กับข้อมูลในชุดตรวจสอบเนื่องจากข้อมูลฝึกน้อยเกินไป จากนั้นเริ่มเพิ่มสูงขึ้น เพราะโมเดลเรียนรู้จากข้อมูลที่หลากหลายขึ้น และใกล้กับเส้น Training Score แสดงว่าโมเดล generalize ได้ดีในชุดตรวจสอบ

**โดยจากกราฟ ช่องว่างระหว่างสองเส้นลดลง** แสดงว่าโมเดลไม่ได้ Overfit หรือ Underfit

* แสดง ROC Curve (Receiver Operating Characteristic Curve) และ AUC (Area Under Curve) ซึ่งเป็นเครื่องมือที่ใช้สำหรับการประเมินประสิทธิภาพของโมเดล Classification

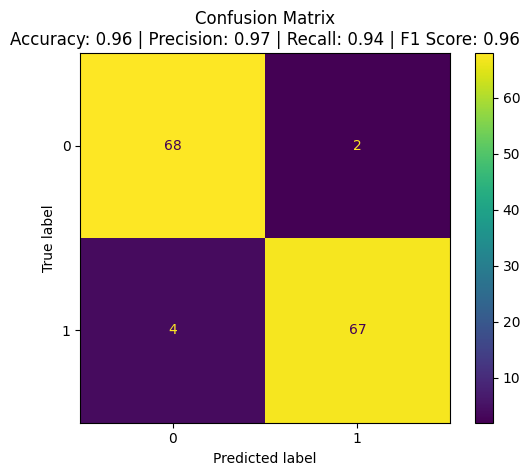


1. **แกน X** แสดงอัตราการทำนายผิด (False Positive Rate) โดยค่าอยู่ระหว่าง 0 ถึง 1
   * คำนวณจาก *FRP* =
2. **แกน Y** แสดงอัตราการทำนายถูก (True Positive Rate) หรือ Recall โดยค่าอยู่ระหว่าง 0 ถึง 1
   * คำนวณจาก *FRP* =
3. **เส้นสีน้ำเงิน** เป็นเส้น ROC Curveแสดงความสัมพันธ์ระหว่าง FPR และ TPR ของโมเดล โดยกราฟโค้งสูงขึ้นและเข้าใกล้มุมบนซ้ายของกราฟอย่างชัดเจน แสดงว่าโมเดลสามารถทำนายกลุ่ม Positive ได้ถูกต้องเกือบทั้งหมด ในขณะที่ยังคงรักษา FPR ให้ต่ำมาก
4. **เส้นประสีแดง** เป็นเส้นมาตรฐานสำหรับการทำนายแบบสุ่ม (Random Guess) ซึ่ง กราฟ ROC Curve สีน้ำเงินห่างจากเส้นทแยงมุมสีแดงอย่างชัดเจน แสดงว่าโมเดลมีความแม่นยำดีกว่าการเดาสุ่ม

**5. ค่า AUC (Area Under the Curve)** พื้นที่ใต้กราฟ ROC ซึ่งบอกถึงประสิทธิภาพโดยรวมของโมเดลซึ่ง

โมเดลนี้มีค่า AUC = 0.99 หมายถึงว่าโมเดลนี้มีประสิทธิภาพสูงมากสามารถแยกระหว่างกลุ่ม Positive และ Negative ได้ดีมาก

* แสดงผลการทำนายในรูปแบบ Confusion Matrix



**แกน X (Predicted)** แสดงค่าผลลัพธ์ที่โมเดลทำนาย (0 หรือ 1)

**แกน Y (Actual)** แสดงค่าจริงที่เกิดขึ้นในชุดข้อมูล (0 หรือ 1)

โดยค่าตัวเลขในแต่ละช่องแสดงจำนวนตัวอย่างที่โมเดลจัดให้อยู่ในกลุ่มนั้น

**ช่อง (True Negative)** – 68 ตัวอย่างที่โมเดลทำนายว่าเป็นคลาส 0 และค่าจริงคือคลาส 0

**ช่อง (False Positive)** – 2 ตัวอย่างที่โมเดลทำนายว่าเป็นคลาส 1 และค่าจริงคือคลาส 0

**ช่อง (False Negative)** – 4 ตัวอย่างที่โมเดลทำนายว่าเป็นคลาส 0 และค่าจริงคือคลาส 1

**ช่อง (True Positives)** – 67 ตัวอย่างที่โมเดลทำนายว่าเป็นคลาส 1 และค่าจริงคือคลาส 1

**Accuracy** (ความแม่นยำของโมเดล): 0.96

**Precision** (ความถูกต้องของการทำนายคลาส 1): 0.97

**Recall** (ความครอบคลุมของการทำนายคลาส 1): 0.94

**F1 Score** (คะแนนเฉลี่ยถ่วงน้ำหนักของ Precision และ Recall): 0.83