# รายงานผลการทดลอง

## ผลการทดลอง

จากการทดลองได้ทำการศึกษากระบวนการ Data Preprocessing โดยใช้ภาษา Python และไลบรารี ได้แก่ `pandas`, `matplotlib`, `seaborn` และ `scikit-learn` เพื่อวิเคราะห์และเตรียมข้อมูลก่อนนำไปใช้ในงาน Machine Learning ซึ่งผลการทดลองมีดังนี้

### 1. Dataset Exploration

สามารถนำเข้าชุดข้อมูล (`data.csv`) ได้สำเร็จ และตรวจสอบข้อมูลเบื้องต้น พบว่าชุดข้อมูลประกอบด้วยหลายคอลัมน์ ได้แก่ Duration, Pulse, Maxpulse และ Calories พร้อมทั้งตรวจสอบจำนวนแถวและคอลัมน์ (Shape) ชนิดข้อมูล (Data Types) และค่าสถิติเบื้องต้น เช่น ค่าเฉลี่ย (Mean), ค่าสูงสุด (Max), ค่าต่ำสุด (Min) และส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)

นอกจากนี้ยังตรวจสอบข้อมูลที่สูญหาย (Missing Values) และข้อมูลซ้ำ (Duplicate Records)
เพื่อเตรียมข้อมูลก่อนเข้าสู่ขั้นตอนการทำความสะอาดข้อมูล

### 2. Data Visualization

ทำการสร้างกราฟ Histogram เพื่อศึกษาการกระจายตัวของข้อมูลแต่ละตัวแปร พบว่าข้อมูลแต่ละคอลัมน์มีการกระจายตัวแตกต่างกัน และสามารถสังเกตค่าผิดปกติ (Outliers) ได้

จากนั้นสร้าง Correlation Heatmap เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปร พบว่าตัวแปร Pulse, Maxpulse และ Calories มีความสัมพันธ์กันในระดับหนึ่ง ซึ่งเป็นข้อมูลที่สามารถนำไปใช้ในการสร้างแบบจำลอง Machine Learning ได้

### 3. Data Cleaning

ดำเนินการทำความสะอาดข้อมูล โดย

- ตรวจสอบและจัดการ Missing Values
- ลบข้อมูลซ้ำ (Duplicate Removal)
- แก้ไขข้อมูลที่ไม่ถูกต้อง (Incorrect Data Correction)
- แปลงชนิดข้อมูล (Data Type Conversion)

หลังจากทำ Data Cleaning แล้ว ชุดข้อมูลมีความสมบูรณ์มากขึ้น ไม่มีข้อมูลซ้ำ
และชนิดข้อมูลมีความเหมาะสมสำหรับการนำไปวิเคราะห์และสร้างแบบจำลอง

### 4. Feature Engineering

ได้ทำการคำนวณค่าทางสถิติ ได้แก่

- Mean
- Median

รวมทั้งแปลงข้อมูลประเภทข้อความให้เป็นตัวเลขด้วยวิธี

- Label Encoding
- One-Hot Encoding

เพื่อให้ข้อมูลสามารถนำไปใช้กับอัลกอริทึม Machine Learning ได้อย่างมีประสิทธิภาพ

## สรุปผลการทดลอง

จากการทดลองพบว่า การทำ Data Preprocessing เป็นขั้นตอนที่สำคัญก่อนการสร้างแบบจำลอง Machine Learning
เนื่องจากช่วยให้ข้อมูลมีคุณภาพมากขึ้น ลดปัญหาที่เกิดจากข้อมูลสูญหาย ข้อมูลซ้ำ และข้อมูลที่ไม่ถูกต้อง
อีกทั้งการสร้างกราฟช่วยให้เข้าใจลักษณะของข้อมูลและความสัมพันธ์ระหว่างตัวแปรได้ดียิ่งขึ้น ส่วนการทำ Feature Engineering
ทำให้ข้อมูลอยู่ในรูปแบบที่อัลกอริทึมสามารถประมวลผลได้
ส่งผลให้การวิเคราะห์ข้อมูลและการสร้างแบบจำลองมีประสิทธิภาพและให้ผลลัพธ์ที่น่าเชื่อถือมากยิ่งขึ้น.