รายงานสรุปโครงการ

ระบบประมวลผล Machine Learning 2 ขั้นตอน เพื่อการจัดการข้อมูลคาร์บอนอย่างยั่งยืนสำหรับ SME

---

1. ภาพรวมของระบบ (Executive Summary)

วิทยานิพนธ์ฉบับนี้เน้นการแก้ปัญหาความไม่น่าเชื่อถือของข้อมูลการใช้พลังงานในโรงงาน SME [cite: 4, 478] โดยเฉพาะในเขตระเบียงเศรษฐกิจภาคตะวันออก (EEC) [cite: 28, 502] ข้อมูลจากเซ็นเซอร์ IoT มักพบปัญหาการขาดหายของข้อมูล (Dropout) ความคลาดเคลื่อน (Drift) หรือการบิดเบือนข้อมูลเพื่อผลประโยชน์ทางการเงิน (Greenwashing) [cite: 39, 526] ระบบนี้จึงถูกพัฒนาขึ้นเพื่อทำหน้าที่เป็น "ผู้ตรวจสอบบัญชีดิจิทัล" ที่ทำงานด้วยปัญญาประดิษฐ์ [cite: 87, 591]

2. ระบบการทำงาน 2 ขั้นตอน (The Two-Stage Pipeline)

ด่านที่ 1: การซ่อมแซมข้อมูล (BiLSTM Imputation)

ใช้โมเดล Bidirectional Long Short-Term Memory (BiLSTM) เพื่อกู้คืนและเติมเต็มข้อมูลส่วนที่ขาดหายไป [cite: 6, 480] โดยอาศัยบริบทของข้อมูลในอดีต อนาคต และตัวแปรแวดล้อม เช่น ยอดการผลิตและสภาพแสงแดด [cite: 233, 806]

ด่านที่ 2: การตรวจสอบความถูกต้อง (Autoencoder Verification)

ใช้โมเดล Autoencoder เรียนรู้ความสัมพันธ์เชิงลึกระหว่างการใช้พลังงานและยอดผลิต (Energy–Production Coupling) [cite: 7, 481] หากพบข้อมูลที่ไม่สอดคล้องกับพฤติกรรมปกติ ระบบจะแจ้งเตือนความผิดปกติทันที [cite: 127, 630]

3. รายละเอียดโครงสร้างไฟล์ในโปรเจกต์

ชื่อไฟล์	หน้าที่และหลักการทำงาน
main.py	ไฟล์หลักที่สั่งการระบบจำลองโรงงานเสมือน (Digital Twin) เพื่อสร้างชุดข้อมูล 1 ปี
config.py	กำหนดค่าพารามิเตอร์ของโรงงาน เช่น พิกัดที่ตั้ง จำนวนเครื่องจักร และสเปกโซลาร์เซลล์
factory_model.py	คำนวณปริมาณไฟฟ้าที่ใช้จริงตามการเดินเครื่องจักรและระบบปรับอากาศ
production_model.py	จำลองตารางการทำงาน กะพนักงาน และเหตุการณ์เครื่องจักรเสียกะทันหัน
solar_model.py	จำลองการผลิตไฟฟ้าจากโซลาร์เซลล์อิงตามข้อมูลสภาพอากาศจริง
fault_injection.py	ตัวสร้างสถานการณ์จำลอง เช่น เซ็นเซอร์เสีย หรือการแอบแก้ตัวเลขยอดผลิต
train_bilstm.py	โค้ดสำหรับฝึกสอน AI ให้รู้วิธีเติมข้อมูลที่หายไปอย่างแม่นยำ
train_autoencoder.py	โค้ดสำหรับฝึกสอน AI ให้จดจำพฤติกรรมปกติเพื่อใช้จับผิดการทุจริต
run_pipeline.py	ระบบทำงานจริงแบบครบวงจรที่ร้อยเรียงทุกโมเดลเข้าด้วยกัน

4. การจำแนกความผิดปกติ (Taxonomy)

ระบบแบ่งระดับความรุนแรงของข้อมูลที่น่าสงสัยตามค่าความผิดพลาดในการสร้างข้อมูลใหม่ (Reconstruction Error):

- **Type I (Marginal):** ข้อมูลเบี่ยงเบนเล็กน้อย อาจเกิดจากช่วงเริ่มเดินเครื่อง
- **Type II (Structural):** ข้อมูลขัดแย้งชัดเจน บ่งชี้ถึงเซ็นเซอร์เสียหรือคีย์ข้อมูลผิด
- **Type III (Critical):** ข้อมูลขัดต่อความเป็นจริงอย่างรุนแรง บ่งชี้ถึงการทุจริตข้อมูล

5. สรุปผลการทดสอบ (Experimental Results)

ผลลัพธ์จากการรันระบบล่าสุด (Integrated Pipeline - Strict Mode):

Precision (ความแม่นยำในการจับผิด)	0.2492
Recall (ความสามารถในการตรวจพบ)	0.0897
F1-Score (ความสมดุล)	0.1320
จำนวนการตรวจพบ (Detected)	618 ชั่วโมง จาก 1,716 ชั่วโมง

หมายเหตุ*:* ผลลัพธ์นี้อ้างอิงตามเกณฑ์การตัดระดับความผิดปกติ *(Threshold)* ที่เปอร์เซ็นไทล์ที่ *99* ตามระเบียบวิธีวิจัยในเอกสารอ้างอิง *[cite: 272, 899]*

อ้างอิงตามมาตรฐาน ISO 14064-3 และนโยบาย Climate Change Act 2024