บทที่ 3
ขั้นตอน และ วิธีการดำเนินงาน

โครงงานเรื่อง การสร้างโมเดลการพยากรณ์ ผลกระทบของปัจจัยทางสภาพภูมิอากาศต่อความเข้มข้นฝุ่นละอองขนาด 2.5 ไมครอน ในเขตอำเภอเมือง จังหวัดเชียงใหม่ โดยใช้การเรียนรู้ของเครื่อง (MACHINE LEARNING) ในบทนี้จะเป็นการวิเคราะห์ข้อมูลด้วยเทคนิค CRISP-DM โดยทำการสร้างโมเดล 2 แบบ คือ 1.โมเดล Random Forest 2. โมเดล LSTM และนำมาเปรียบเทียบประสิทธิภาพของโมเดลทั้งสอง โดยใช้ค่า RMSE และ MAE เป็นตัวเปรียบเทียบ จากนั้นทำการนำโมเดลที่ได้ไปทำการพัฒนาเว็บไซต์ สำหรับทำการทำนาย และ แสดงผล

3.1 ขั้นตอน และ วิธีการวิเคราะห์ข้อมูลด้วย CRISP-DM

3.1.1 กระบวนการศึกษาทำความเข้าใจธุรกิจ (Business Understanding)

3.1.2 การทำความเข้าใจข้อมูล (Data Understanding)

3.1.3 การเตรียมข้อมูล (Data Preparation)

3.1.3.1 การสกัดข้อมูลที่ต้องการ (Extract)

3.1.3.2 การเปลี่ยนรูปข้อมูลให้เหมาะสม (Transform)

3.1.3.3 การโหลดข้อมูล (Load)

3.1.3.3.1 การโหลดข้อมูลเข้าสู่โปรแกรม RapidMiner

3.1.3.3.2 การโหลดข้อมูลเข้าสู่เว็บไซต์ Google Drive

3.1.4 การสร้างโมเดล (Modeling)

3.1.4.1 การสร้างโมเดลการทำนายโดยใช้โมเดล LSTM

3.1.4.2 การสร้างโมเดลการทำนายโดยใช้โมเดล Random Forest

3.1.5 การวัดประสิทธิภาพของโมเดล (Evaluation)

3.1.5.1 การวัดประสิทธิภาพของโมเดล LSTM

3.1.5.2 การวัดประสิทธิภาพของโมเดล Random Forest

3.1.5.3 การเปรียบเทียบประสิทธิภาพระหว่าง โมเดล LSTM กับโมเดล Random Forest

3.1.6 การนำโมเดลไปใช้งานจริง (Deployment)

3.1.6.1 การสร้างหน้าเว็บไซต์สำหรับแสดงผล

3.1.6.2 การสร้างส่วนการทำนายผลในหน้าเว็บไซต์

3.1.6.2 การอัปโหลดเว็บไซต์และเผยแพร่

3.1.1 กระบวนการศึกษาทำความเข้าใจธุรกิจ (Business Understanding)

เป็นขั้นตอน แรกของกระบวนการ ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวมเป็น ขั้นตอนแรกของกระบวนการ ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม ผู้ วิเคราะห์ข้อมูลทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็นนี้ ข้อมูลจะถูกจัดเก็บโดยหน่วยงานราชการ ซึ่งมีทั้งหมด 2 หน่วยงาน 1. กรมควบคุมมลพิษ 2. ศูนย์อุตุนิยมวิทยาภาคเหนือ ซึ่งข้อมูลจะถูกจัดเก็บไว้จากทั้งประเทศโดยแยกออกเป็นหน่วยพื้นที่ต่าง ๆ

3.1.2 การทำความเข้าใจข้อมูล (Data Understanding)

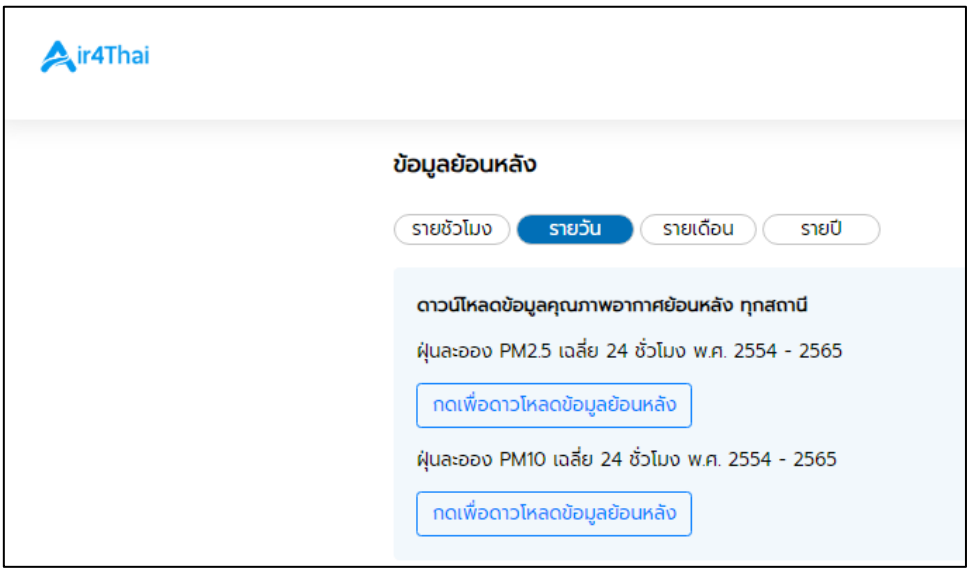
ขั้นตอนการจัดเก็บและรวบรวม ข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูล ทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดรับกับวัตถุประสงค์ที่กำหนดไว้

ผู้วิเคราะห์ทำการรวบรวมข้อมูล เพื่อทำการวิเคราะห์ ข้อมูลความกดอากาศ ข้อมูลความชื้น ข้อมูลความเร็วลม ข้อมูลปริมาณน้ำฝน ข้อมูลอุณหภูมิ ทำการรวบรวมมาจากเว็บไซต์ของ ศูนย์อุตุนิยมวิทยาภาคเหนือ เข้าถึงได้จาก http://www.cmmet.tmd.go.th ซึ่งเป็นข้อมูลเปิดเผยต่อสาธารณะ

ภาพที่ 3.1 ศูนย์อุตุนิยมวิทยาภาคเหนือ

(ที่มา : http://www.cmmet.tmd.go.th)

ข้อมูล PM2.5 ทำการรวบรวมมาจากเว็บไซต์ของ กรมควบคุมมลพิษ เข้าถึงได้จาก http://air4thai.pcd.go.th ซึ่งเป็นข้อมูลที่เปิดเผยต่อสาธารณะ



ภาพที่ 3.2 หน้าเว็บไซต์ของกรมควบคุมมลพิษ

(ที่มา : http://air4thai.pcd.go.th)

3.1.3 การเตรียมข้อมูล (Data Preparation)

3.1.3.1 การสกัดข้อมูลที่ต้องการ (Extract)

ดาวน์โหลดข้อมูลจากแหล่งข้อมูลโดยที่ ข้อมูลความกดอากาศ ข้อมูลความชื้น ข้อมูลความเร็วลม ข้อมูลปริมาณน้ำฝน ข้อมูลอุณหภูมิ ทำการรวบรวมมาจากเว็บไซต์ของ ศูนย์อุตุนิยมวิทยาภาคเหนือ เข้าถึงได้จาก http://www.cmmet.tmd.go.th

ข้อมูล PM2.5 ทำการรวบรวมมาจากเว็บไซต์ของ กรมควบคุมมลพิษ เข้าถึงได้จาก http://air4thai.pcd.go.th

3.1.3.2 การเปลี่ยนรูปข้อมูลให้เหมาะสม (Transform)

ทำการแปลงข้อมูลให้พร้อมสำหรับการนำไปใช้งานในการเทรนโมเดล โดยจะทำการรวมข้อมูลที่ต้องการไว้ในไฟล์เดียวกัน โดยจะทำการคัดแยก ข้อมูลความกดอากาศ ข้อมูลความชื้น ข้อมูลความเร็วลม ข้อมูลปริมาณน้ำฝน ข้อมูลอุณหภูมิ และ PM2.5 ในช่วงปี พ.ศ. 2561 – 2565 รวมไว้ในไฟล์เดียวกัน

ทำความสะอาดข้อมูล โดยทำการทำการแปลงข้อมูลให้อยู่ในรูปแบบที่ต้องการ จัดการค่าว่างของข้อมูล และ ทำการปรับปรุงข้อมูลให้อยู่ในมาตรฐานเดียวกัน

3.1.3.3 การโหลดข้อมูล (Load)

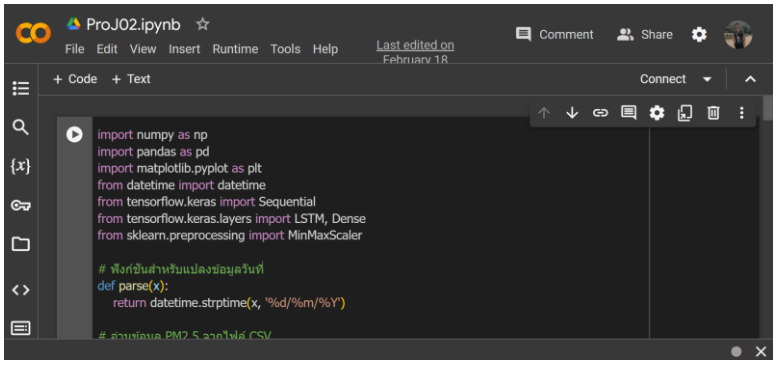
การนำเข้าข้อมูลไปยัง Google Drive เพื่อเตรียมใช้ในการเทรนโมเดล LSTM

การนำเข้าข้อมูลไปยัง Rapid Miner เพื่อเตรียมใช้ในการเทรนโมเดล Random Forest

3.1.4 การสร้างโมเดล (Modeling)

3.1.4.1 การสร้างโมเดลการทำนายโดยใช้โมเดล LSTM

โมเดล LSTM ทำการสร้างและทดสอบโดยใช้เว็บไซต์ Google Colab

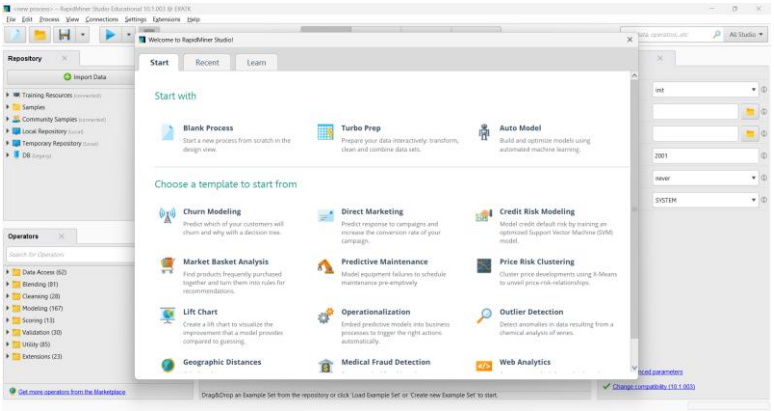


ภาพที่ 3.3 หน้าต่างเว็บไซต์ Google Colab

1) ติดตั้ง Library ที่จำเป็นสำหรับการใช้ในการสร้างโมเดล และ Library ที่จำเป็นอื่น ๆ
2) การอ่านและนำเข้าข้อมูลสำหรับใช้ในการทำนาย โดยนำเข้าข้อมูลจาก Google Drive
3) ทำการตั้งค่าโมเดล และ กำหนดโครงสร้าง Neural Network
4) แบ่งข้อมูลออกเป็น 2 ชุดเพื่อนำไปสร้างชุดข้อมูลสอน และ ชุดข้อมูลทดสอบ โดยแบ่งอัตราส่วน 0.8:0.2 (อ้างอิงจาก)
5) ทำการสร้างโมเดลและเทรนด้วยข้อมูลชุดตัวอย่าง

3.1.4.2 การสร้างโมเดลการทำนายโดยใช้โมเดล Random Forest

โมเดล Random Forest จะทำการสร้างและวัดผลในโปรแกรม RapidMiner



ภาพที่ 3.4 หน้าแอปพลิเคชั่น RapidMiner

1) เลือกชุดข้อมูลที่นำเข้าสู่โปรแกรม Rapid Miner
2) แบ่งข้อมูลออกเป็น 2 ชุดเพื่อนำไปสร้างชุดข้อมูลสอน และ ชุดข้อมูลทดสอบ โดยแบ่งอัตราส่วน 0.8:0.2 (อ้างอิงจาก)
3) สร้างโมเดล Random Forest โดยใช้ Parameter Random forest

3.1.5 การวัดประสิทธิภาพของโมเดล (Evaluation)

3.1.5.1 การวัดประสิทธิภาพของโมเดล LSTM

1) วัดผลประสิทธิภาพ โดยคำนวณค่า RMSE และ MAE

3.1.5.2 การวัดประสิทธิภาพของโมเดล Random Forest

1) วัดผลประสิทธิภาพ โดยคำนวณค่า RMSE และ MAE

3.1.5.3 เปรียบเทียบประสิทธิภาพระหว่าง โมเดล LSTM กับ โมเดล Random Forest

โดยทำการเปรียบเทียบ 2 ส่วน

1) เปรียบเทียบกับค่าที่ยอมได้ ซึ่งค่า RMSE ต้องน้อยกว่า 7.00 และค่า MAE ต้องน้อยกว่า 27.00 กฤติกา ทิพย์คำมี และ คณะ (2566)
2) เปรียบเทียบระหว่างโมเดล LSTM กับ โมเดล Random Forest โดยทำการเลือกโมเดลที่มีค่าน่าพึงพอใจที่สุด

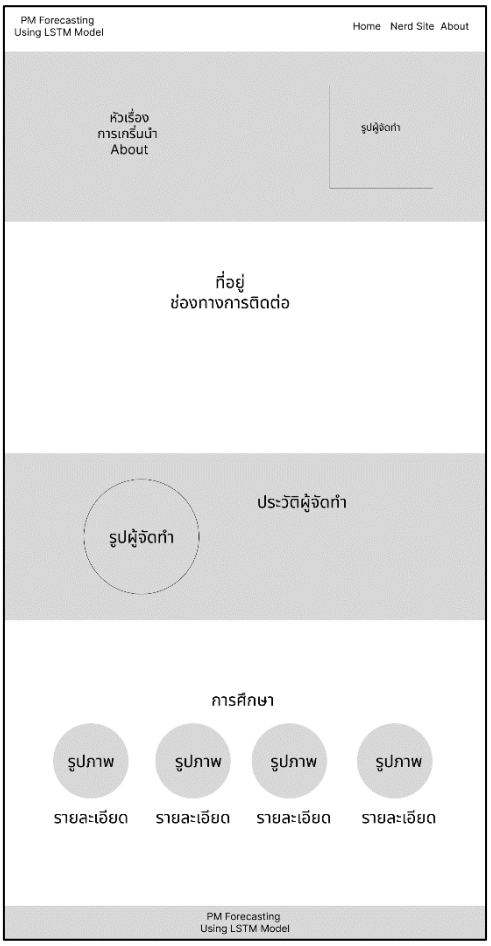
ตารางที่ 3.1 ตารางเปรียบเทียบประสิทธิภาพระหว่างโมเดล

	MAE	RMSE
Random Forest		
LSTM		

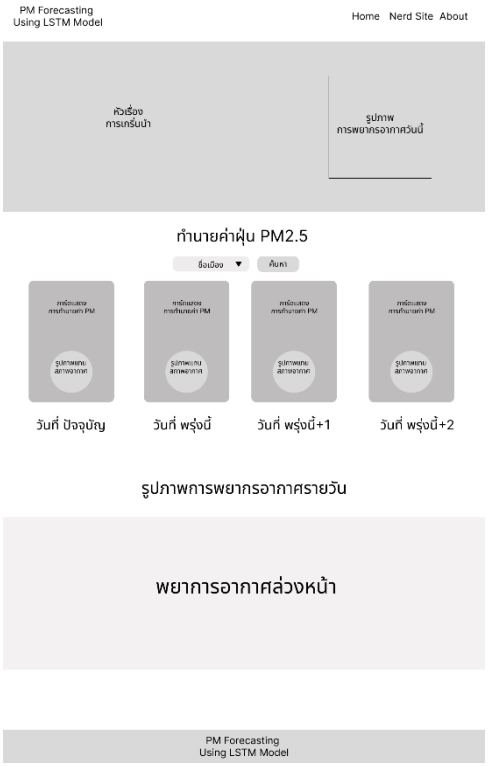
3.1.6 การนำโมเดลไปใช้งานจริง (Deployment)

3.1.6.1 การสร้างหน้าเว็บไซต์สำหรับแสดงผล

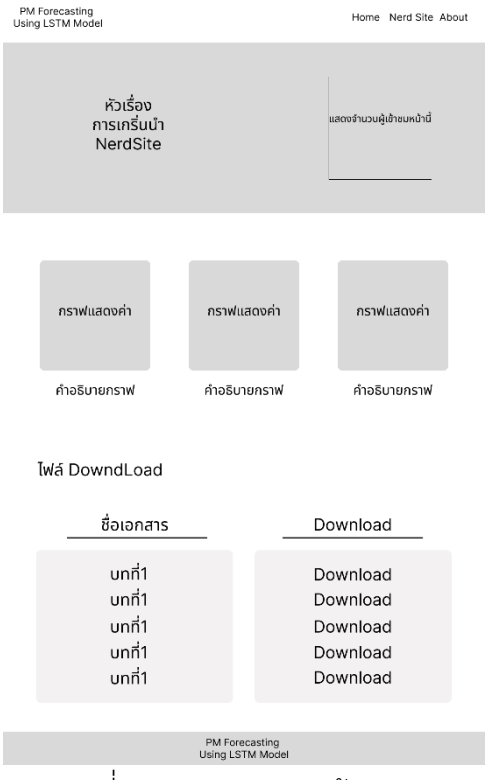
1) สร้างหน้าเว็บไซต์ โดยใช้โปรแกรม Nice Page ซึ่งเป็นโปรแกรมช่วยในการสร้างหน้าเพจ HTML และ CSS เบื้องต้นโดยทำการสร้างตาม Wireframe ที่ได้ออกแบบไว้



ภาพที่ 3.5 Wireframe หน้า Index



ภาพที่ 3.6 Wireframe หน้า NerdSite



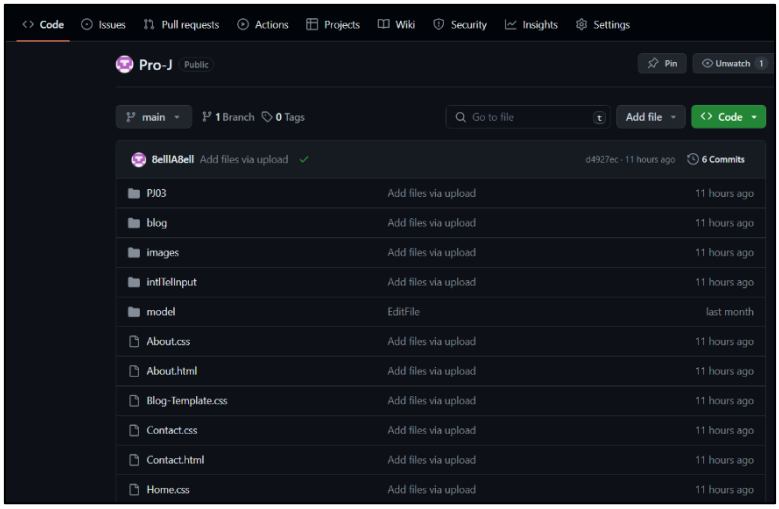
ภาพที่ 3.7 Wireframe หน้า About

3.1.6.2 การสร้างส่วนการทำนายผลในหน้าเว็บไซต์

1) สร้างและทดสอบการใช้โมเดลในหน้าเว็บไซต์ โดยใช้ภาษา Java Script และ ทำการนำค่าทำนายที่ได้ ทำการแสดงผลผ่านทางหน้าเว็บเพจ

3.1.6.2 การอัปโหลดเว็บไซต์และเผยแพร่

1) อัปโหลดไฟล์หน้าเว็บขึ้นสู่เว็บไซต์ GitHub เป็นบริการ Web Hosting สำหรับเว็บไซต์ขนาดเล็ก เพื่อให้เว็บไซต์สามารถเข้าถึงได้จากอินเตอร์เน็ต



ภาพที่ 3.8 หน้าต่างโปรแกรม GitHub Desktop