โครงการ

การเรียนการสอนเพื่อเสริมประสบการณ์

ชื่อโครงการ การประมาณค่าความหนาแน่นของ PM 2.5 ด้วยการเรียนรู้ของเครื่อง
A Density Estimation of PM 2.5 using Machine Learning

ชื่อนิสิต นายศีลวัต หมื่นสุวรรณ 6033539123

ภาควิชา คณิตศาสตร์และวิทยาการคอมพิวเตอร์
สาขาวิชา คณิตศาสตร์

ปีการศึกษา 2563

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

การประมาณค่าความหนาแน่นของ PM 2.5 ด้วยการเรียนรู้ของเครื่อง

นายศีลวัต หมื่นสุวรรณ

โครงงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2563
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A Density Estimation of PM 2.5 using Machine Learning

Seenlawat Muensuwan

A Project Submitted in Partial Fulfillment of the Requirements

for the Degree of Bachelor of Science Program in Mathematics

Department of Mathematics and Computer Science

Faculty of Science

Chulalongkorn University

Academic Year 2020

Copyright of Chulalongkorn University

หัวข้อโครงงาน	การประมาณค่าความหนาแน่นของ PM 2.5 ด้วยการเรียนรู้ของเครื่อง
โดย	นายศีลวัต หมื่นสุวรรณ
สาขาวิชา	คณิตศาสตร์
อาจารย์ที่ปรึกษาโครงงาน	รองศาสตราจารย์ ดร. ภัทรสินี ภัทรโกศล

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับโครงงานฉบับนี้เป็นส่วนหนึ่ง ของการศึกษาตามหลักสูตรปริญญาบัณฑิต ในรายวิชา 2301499 โครงงานวิทยาศาสตร์ (Senior Project)

..... หัวหน้าภาควิชาคณิตศาสตร์

(ศาสตราจารย์ ดร. กฤษณะ เนียมมณี) และวิทยาการคอมพิวเตอร์

คณะกรรมการสอบโครงงาน

..... อาจารย์ที่ปรึกษาโครงงาน

(รองศาสตราจารย์ ดร. ภัทรสินี ภัทรโกศล)

..... กรรมการ

(ผศ. ดร. กีรติ ศรีอมร)

..... กรรมการ

(ผศ. ดร. ธีระเดช กิตติภัสสร)

นายศีลวัต หมื่นสุวรรณ: การประมาณค่าความหนาแน่นของ PM 2.5 ด้วยการเรียนรู้ของเครื่อง. (A Density Estimation of PM 2.5 using Machine Learning) อ.ที่ปรึกษาโครงงาน: รศ.ดร. ภัทรสินี ภัทรโกศล, 51 หน้า.

โครงการวิจัยเรื่อง “การประมาณค่าความหนาแน่นของ PM 2.5 ด้วยการเรียนรู้ของเครื่อง” มีวัตถุประสงค์ คือ กำหนดปัจจัยที่มีผลต่อการเกิด PM 2.5 และ พัฒนาโมเดลความสัมพันธ์ของปัจจัยที่พิจารณากับปริมาณ PM 2.5 ที่วัดได้ จุดมุ่งหมายของงานวิจัยเพื่อให้ทราบว่าปัจจัยใดบ้างที่ส่งผลต่อ PM 2.5 ที่ทำให้ระดับความอันตรายของ PM 2.5 มีค่าสูง หรือต่ำ ขอบเขตการวิจัยภายใต้ข้อมูลเชิงประวัติของค่า PM 2.5 ที่จะนำมาพิจารณาจะอยู่ในช่วงปีพุทธศักราช 2563 เท่านั้น และแหล่งข้อมูลที่ใช้ คือ สำนักงานสถิติแห่งชาติ กรมอุตุนิยมวิทยา กรมควบคุมมลพิษ เว็บไซต์ Thai Meteorlogical Department เว็บไซต์ Air Quality Historical Data Platform และ Google Earth วิธีการวิจัยซึ่งมีวัตถุประสงค์ในการวิเคราะห์ข้อมูล PM 2.5 เพื่อนำมาหาโมเดลการคาดคะเนที่ดีที่สุดของปริมาณความเป็นมลพิษของ PM 2.5 ด้วยโปรแกรม RapidMiner Studio ผลการวิจัยที่ได้แสดงให้เห็นว่าปัจจัยที่มีผลต่อ PM 2.5 ทางสนับสนุน หรือทำให้ระดับความอันตรายสูงขึ้น คือ จำนวนประชากรหญิงของจังหวัด ปริมาณหยาดน้ำฟ้า จำนวนประชากรหญิงของอำเภอ ความกดอากาศที่ระดับน้ำทะเล และความสูงจากระดับน้ำทะเล ส่วนปัจจัยที่ส่งผลในทางตรงข้าม หรือ ทำให้ค่าความอันตรายลดลง คือ จำนวนประชากรรวมของอำเภอ จำนวนประชากรชายของอำเภอ ฤดูกาล อุณหภูมิ และความชื้น

ภาควิชา คณิตศาสตร์และวิทยาการคอมพิวเตอร์ ลายมือชื่อนิสิต ศีลวัต

สาขาวิชา คณิตศาสตร์ ลายมือชื่อ อ.ที่ปรึกษาโครงงาน ภัทรสินี ภัทรโกศล

ปีการศึกษา 2563

6033539123: MAJOR MATHEMATICS

KEYWORDS : PM 2.5 / MACHINE LEARNING / RAPIDMINER STUDIO

SEENLAWAT MUENSUWAN: A DENSITY ESTIMATION OF PM 2.5 USING MACHINE LEARNING. ADVISOR : ASSOC. PROF. PATTARASINEE BHATTARAKOSOL, Ph.D., 51 pp.

The topic of this research is "A Density Estimation of PM 2.5 using Machine Learning". The objectives of this research are to define important factors for PM 2.5 and to develop relationships between considered factors and PM 2.5. The aim of this research is to determine factors those affect the risk-level of PM 2.5 as high or low. The research scopes on the the historical data of PM 2.5 values in the year 2020 only, and the data sources are the National Statistical Office, Meteorological Department, Pollution Control Department, Thai Meteorlogical Department website, Air Quality Historical Data Platform website, and Google Earth. The research method aims to analyze the collected data of PM 2.5 to find the best volume-prediction model of the PM 2.5 pollution using RapidMiner Studio. The results have shown that factors affecting PM 2.5 those support or push the danger level are female populations of the province, precipitation, female populations of the district, air pressure at sea level, and metres above sea level. In contrast, the total population of the district, male populations of the district, season, temperature, and humidity are all factors that can decrease the hazard level of PM 2.5.

Department :Mathematics and Computer Science Student's Signature ศีลาวัต

Field of Study : Mathematics Advisor's Signature P Bhattarakosol

Academic Year : 2020

กิตติกรรมประกาศ

การวิจัยในชั้นเรียนในหัวข้อเรื่อง “การประมาณค่าความหนาแน่นของ PM 2.5 ด้วยการเรียนรู้ของเครื่อง” ได้รับการอนุเคราะห์จากบุคคลหลายท่าน ซึ่งไม่อาจจะนำมากล่าวได้ทั้งหมด ซึ่งผู้วิจัยใคร่ขอกราบขอบพระคุณ คือ รศ.ดร. ภัทรสินี ภัทรโกศล ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ ซึ่งท่านเป็นอาจารย์ที่ปรึกษาในการวิจัยครั้งนี้ อาจารย์ได้คอยให้คำปรึกษา คำแนะนำ ช่วยตรวจทาน และแก้ไขข้อบกพร่องต่าง ๆ ของโครงงานนี้ และการสนับสนุนอย่างเต็มที่จากกรมอุตุนิยมวิทยา ในด้านข้อมูลทางอุตุนิยมวิทยา และกรมควบคุมมลพิษ ในด้านข้อมูล PM 2.5 รวมถึงขอบคุณแฟนที่ให้กำลังใจเสมอ และคอยเตือนให้ทำงานให้ส่งตามกำหนด ซึ่งทำให้งานวิจัยนี้สำเร็จลุล่วงไปด้วยดี ข้าพเจ้าจึงใคร่ขอขอบพระคุณเป็นอย่างยิ่งสำหรับความช่วยเหลือในทุก ๆ ด้าน ที่ได้รับจากทุกบุคคล และหวังว่าผลการวิจัยนี้จะเป็นประโยชน์ในการประมาณค่าความหนาแน่นของ PM 2.5 ได้ต่อไป

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฌ

สารบัญภาพ ..... ญ

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและเหตุผลการวิจัย 1

1.2 วัตถุประสงค์ของการวิจัย 2

1.3 ขอบเขตการวิจัย 2

1.4 ขั้นตอนการวิจัย 2

1.5 ประโยชน์ที่คาดว่าจะได้รับ 3

1.6 โครงสร้างของรายงาน 4

บทที่ 2 งานวิจัยที่เกี่ยวข้อง 5

2.1 การเรียนรู้ของเครื่อง (Machine Learning) 5

2.2 Particulate Matter 2.5 (PM 2.5) 11

2.3 RapidMiner Studio 14

บทที่ 3 วิธีการวิจัย 17

3.1 วิเคราะห์ข้อมูล PM 2.5 และการคาดคะเน 17

3.2 เก็บรวบรวมข้อมูล 17

3.3 นำข้อมูลมาวิเคราะห์ด้วยโปรแกรม 18

3.4 การเลือกโมเดลที่ที่ดีสุด 25

บทที่ 4 ผลการวิจัย 29
4.1 กรณีที่เราไม่ต้องใส่หน้ากากอนามัย (0) 29
4.2 กรณีที่เราควรใส่หน้ากากอนามัย (1) 30
บทที่ 5 ข้อสรุปและข้อเสนอแนะ 31
5.1 ข้อสรุป 31
5.2 ข้อเสนอแนะ 32
รายการอ้างอิง 33
แบบเสนอหัวข้อโครงงาน รายวิชา 2301399 Project Proposal ปีการศึกษา 2563 36
ประวัติผู้เขียน 41

สารบัญตาราง

หน้า

ตารางที่ 1.1 ระยะเวลาของขั้นตอนการวิจัย 3
ตารางที่ 2.1 อัลกอริทึมแบบต่าง ๆ 8
ตารางที่ 2.2 เกณฑ์ของดัชนีคุณภาพอากาศของประเทศไทย 12

สารบัญภาพ

หน้า

รูปที่ 2.1 หน้าต่างโปรแกรม RapidMiner Studio 15
รูปที่ 2.2 หน้าต่าง Operators 15
รูปที่ 2.3 หน้าต่าง Process และ Parameters 16
รูปที่ 2.4 หน้าต่าง Help 16
รูปที่ 3.1 Excel ตารางข้อมูล 18
รูปที่ 3.2 ขั้นตอนที่ 1 19
รูปที่ 3.3 ขั้นตอนที่ 2 19
รูปที่ 3.4 ขั้นตอนที่ 3 20
รูปที่ 3.5 ขั้นตอนที่ 4 20
รูปที่ 3.6 ขั้นตอนที่ 5 21
รูปที่ 3.7 ขั้นตอนที่ 6 21
รูปที่ 3.8 ขั้นตอนที่ 7 22
รูปที่ 3.9 ขั้นตอนที่ 8 22
รูปที่ 3.10 ขั้นตอนที่ 9 23
รูปที่ 3.11 ขั้นตอนที่ 10 23
รูปที่ 3.12 ขั้นตอนที่ 11 24
รูปที่ 3.13 ขั้นตอนที่ 12 24
รูปที่ 3.14 Performance ของ Naïve Bayes 25
รูปที่ 3.15 Performance ของ Generalized Lenear Model 25
รูปที่ 3.16 Performance ของ Logistic Regression 26
รูปที่ 3.17 Performance ของ Fast Large Margin 26
รูปที่ 3.18 Performance ของ Deep Learning 27
รูปที่ 3.19 Performance ของ Decision Tree 27
รูปที่ 3.20 Performance ของ Random Forest 28
รูปที่ 3.21 Performance ของ Gradient Boosted Trees 28
รูปที่ 4.1 ปัจจัยที่สำคัญต่อกรณีที่เราไม่ต้องใส่หน้ากากอนามัย (0) 29
รูปที่ 4.2 ปัจจัยที่สำคัญต่อกรณีที่เราควรใส่หน้ากากอนามัย (1) 30

บทที่ 1
บทนำ

1.1 ความเป็นมาและเหตุผลการวิจัย

เนื่องด้วยปัจจุบัน PM 2.5 เป็นปัญหาอย่างหนึ่งของประเทศไทยที่นับได้ว่าเป็นหนึ่งในปัญหาร้ายแรงในช่วงนี้ โดย PM ย่อมาจาก Particulate Matters คือ ค่ามาตรฐานของฝุ่นละอองที่ลอยในอากาศ ส่วน 2.5 คือ หน่วยวัดเป็นไมครอน ซึ่งเล็กกว่าเส้นผมมนุษย์หลายสิบเท่า จนไม่สามารถมองเห็นได้ด้วยตาเปล่า PM 2.5 เกิดจากการเผาไหม้ทั้งจากยานพาหนะ การเผาวัสดุการเกษตร ไฟป่า และกระบวนการอุตสาหกรรม PM 2.5 สามารถลอยละล่องอยู่ในอากาศเวียนไปวนมา เมื่อมนุษย์หายใจรับ PM 2.5 ผ่านเข้าจมูก ผ่านปาก หรือแม้แต่มาสัมผัสกับร่างกาย ทั้งเข้าตา ทั้งโดนใบหน้า มันสามารถเข้าไปถึงถุงลมในปอดได้ เป็นผลทำให้เกิดโรคในระบบทางเดินหายใจ และโรคปอดต่าง ๆ หากได้รับในปริมาณมากหรือเป็นระยะเวลานานจะสะสมในเนื้อเยื่อปอด ทำให้การทำงานของปอดเสื่อมประสิทธิภาพลง ทำให้หลอดลมอักเสบ มีอาการหอบหืด หรืออาจส่งผลอันตรายถึงชีวิต

PM 2.5 เริ่มเป็นที่รู้จักในคนไทยมากขึ้นในช่วง ธันวาคม 2560 เนื่องจากในช่วงนั้น PM 2.5 ได้มีปริมาณสูงขึ้นเป็นอย่างมากและส่งผลให้หน้ากากอนามัยมีราคาสูงขึ้นอย่างที่ไม่เคยเป็นมาก่อน รวมถึงทำให้เกิดค่าเสียโอกาสจากประเด็นด้านสุขภาพและการท่องเที่ยว รวมถึงค่าเสียโอกาสของภาคธุรกิจที่อาจได้รับผลกระทบจากการที่ประชาชนมีการปรับเปลี่ยนพฤติกรรม จึงก่อให้เกิดผลกระทบทางเศรษฐกิจ อีกทั้งประชาชนบางกลุ่มยังมีภาระในค่าใช้จ่ายด้านสุขภาพ และนักท่องเที่ยวต่างชาติเลือกที่จะปรับกิจกรรมการท่องเที่ยวของตน โดยอาจมีบางส่วนที่หลีกเลี่ยงการเดินทางเข้ามายังพื้นที่กรุงเทพมหานครและปริมณฑล

ด้วยเหตุนี้เอง รัฐบาลได้ออกมาตรการเพื่อใช้แก้ปัญหาที่เกิดขึ้น เช่น แผนปฏิบัติการขับเคลื่อนวาระแห่งชาติ การแก้ไขปัญหามลพิษด้านฝุ่นละอองขนาดเล็ก PM 2.5 การติดตามตรวจสอบปริมาณฝุ่น PM2.5 และการดำเนินงานของกรมควบคุมมลพิษ มาตรการเพิ่มประสิทธิภาพในการบริหารจัดการเชิงพื้นที่ มาตรการและการป้องกันและลดการเกิดมลพิษที่ต้นทาง และมาตรการเพิ่มประสิทธิภาพการบริหารจัดการมลพิษ

ผู้วิจัยจึงเล็งเห็นความสำคัญของเรื่องนี้และต้องการที่จะรู้ว่า ปัจจัยใดมีผลต่อการเกิด PM 2.5 และต้องการที่จะพัฒนาให้เกิดเป็นโมเดลความสัมพันธ์ของปัจจัยที่พิจารณากับปริมาณ PM 2.5 ที่วัดได้ เพราะว่าผลกระทบของ PM 2.5 นั้นมีมากมาย ซึ่งหากสามารถคาดการณ์ปริมาณของ PM 2.5 ในอนาคตภายใต้ปัจจัยที่สามารถวัดได้ หน่วยงานที่เกี่ยวข้องต่าง ๆ ก็จะสามารถรับมือ แก้ไข และเตรียมพร้อมเพื่อลดความเสียหายที่จะเกิดขึ้นอันเกิดจาก PM 2.5 ได้

1.2 วัตถุประสงค์ของการวิจัย

1. กำหนดปัจจัยที่มีผลต่อการเกิด PM 2.5
2. พัฒนาโมเดลความสัมพันธ์ของปัจจัยที่พิจารณากับปริมาณ PM 2.5 ที่วัดได้

1.3 ขอบเขตการวิจัย

1. ค่า PM 2.5 ที่จะนำมาพิจารณาจะอยู่ในช่วงปีพุทธศักราช 2563 เท่านั้น
2. แหล่งข้อมูลที่ใช้ คือ สำนักงานสถิติแห่งชาติ กรมอุตุนิยมวิทยา กรมควบคุมมลพิษ เว็บไซต์ Thai Meteorlogical Department เว็บไซต์ Air Quality Historical Data Platform และ Google Earth

1.4 ขั้นตอนการวิจัย

การวิจัยเพื่อการประมาณค่าความหนาแน่นของ PM 2.5 ด้วยการเรียนรู้ของเครื่อง มีขั้นตอนการดำเนินการดังต่อไปนี้

1. ศึกษาหาความรู้เกี่ยวกับ PM 2.5 และปัจจัยต่าง ๆ ที่ส่งผลกระทบต่อ PM 2.5
2. ศึกษาวิธีใช้โปรแกรม RapidMiner Studio
3. เก็บรวบรวมข้อมูล PM 2.5 จากแหล่งข้อมูลต่าง ๆ
4. หาปัจจัยที่มีผลต่อการเกิด PM 2.5 โดยดูจากความสัมพันธ์ของค่านั้น ๆ กับ PM 2.5 โดยพิจารณาร่วมกับข้อมูลจากบทความวิจัย
5. หาโมเดลความสัมพันธ์เพื่อแสดงให้เห็นได้ว่า ปริมาณการเกิด PM 2.5 สามารถประมาณค่าได้จากปัจจัยต่าง ๆ ได้อย่างไร จากการใช้โปรแกรม RapidMiner Studio
6. สรุปผล และเขียนรายงาน

ตารางที่ 1.1 ระยะเวลาของขั้นตอนการวิจัย

ขั้นตอน	เดือน									
	พ.ค.	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.
1.ศึกษาหาความรู้เกี่ยวกับ PM 2.5 และปัจจัยต่าง ๆ ที่ส่งผลกระทบต่อ PM 2.5	■	■								
2.ศึกษาวิธีใช้โปรแกรม RapidMiner Studio		■								
3.เก็บรวบรวมข้อมูล PM 2.5 จากแหล่งข้อมูลต่าง ๆ		■	■	■	■					
4.หาปัจจัยที่มีผลต่อการเกิด PM 2.5 โดยดูจากความสัมพันธ์ของค่านั้น ๆ กับ PM 2.5 โดยพิจารณาร่วมกับข้อมูลจากบทความวิจัย					■	■	■			
5.หาโมเดลความสัมพันธ์เพื่อแสดงให้เห็นได้ว่าปริมาณการเกิด PM 2.5 สามารถประมาณค่าได้จากปัจจัยต่าง ๆ ได้อย่างไร จากการใช้โปรแกรม RapidMiner Studio							■	■		
6.สรุปผล และเขียนรายงาน								■	■	■

1.5 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่ได้รับจากการวิจัยในครั้งนี้มีดังนี้

1.5.1 สำหรับผู้ที่อ่านงานวิจัยนี้

1.5.1.1 ได้รู้จักกับ PM 2.5 มากขึ้น

1.5.1.2 รู้ว่าปัจจัยใดบ้างที่มีผลต่อ PM 2.5

1.5.1.3 ใช้โมเดลช่วยในการคาดการณ์ค่า PM 2.5 ในอนาคต เพื่อลดความเสียหายที่จะเกิดขึ้น

1.5.1.4 สามารถป้องกันตัวเองจาก PM 2.5

1.5.2 สำหรับผู้ทำงานวิจัยนี้

1.5.2.1 พัฒนาความสามารถในการคิด วิเคราะห์

1.5.2.2 เรียนรู้การวิเคราะห์ข้อมูล ปัจจัยใดที่ส่งผลกระทบกับประเด็นที่เราต้องการ

1.5.2.3 เรียนรู้การใช้งานโปรแกรม RapidMiner Studio

1.5.2.4 เรียนรู้วิธีการติดต่อกับหน่วยงานราชการ

1.5.2.5 มีประสบการณ์ในการทำงานจริง

1.6 โครงสร้างของรายงาน

บทที่ 2 จะกล่าวถึงงานวิจัยที่เกี่ยวข้องกับการประมาณค่าความหนาแน่นของ PM 2.5 ด้วยการเรียนรู้ของเครื่อง

บทที่ 3 จะกล่าวถึงวิธีการวิจัย ซึ่งประกอบไปด้วยวิเคราะห์ข้อมูล PM 2.5 และทำการคาดคะเน เก็บรวบรวมข้อมูล นำข้อมูลมาวิเคราะห์ด้วยโปรแกรม เลือกโมเดลที่ดีที่สุด

บทที่ 4 จะกล่าวถึงผลการวิจัย ซึ่งได้ว่าปัจจัยที่มีผลต่อ PM 2.5 ทางสนับสนุน คือ จำนวนประชากรหญิงของจังหวัด ปริมาณหยาดน้ำฟ้า จำนวนประชากรหญิงของอำเภอ ความกดอากาศที่ระดับน้ำทะเล และความสูงจากระดับน้ำทะเล ส่วนทางตรงข้าม คือ จำนวนประชากรรวมของอำเภอ จำนวนประชากรชายของอำเภอ ฤดูกาล อุณหภูมิ และความชื้น

บทที่ 5 จะกล่าวถึงข้อสรุป และข้อเสนอแนะ

บทที่ 2
งานวิจัยที่เกี่ยวข้อง

บทนี้จะกล่าวถึงงานวิจัยที่เกี่ยวข้องกับระบบการประมาณค่าความหนาแน่นของ PM 2.5 ด้วยการเรียนรู้ของเครื่อง

2.1 การเรียนรู้ของเครื่อง (Machine Learning)

2.1.1 ความหมายของการเรียนรู้ของเครื่อง

การเรียนรู้ของเครื่อง (Machine Learning) หมายถึง การให้เครื่องเรียนรู้งานใดงานหนึ่ง (Task) จากตัวอย่าง (Sample) หรือประสบการณ์ (Experience) จำนวนหนึ่งเพื่อให้ทำงานนั้น ๆ ได้อย่างมีประสิทธิภาพ อีกทั้งยังสามารถเพิ่มประสิทธิภาพ (Performance) ได้จากการเรียนรู้จากตัวอย่างหรือประสบการณ์ที่เพิ่มขึ้นได้

การเรียนรู้ของเครื่อง (Machine Learning) เป็นเครื่องมือหนึ่งของปัญญาประดิษฐ์ที่มุ่งเน้นในการใช้ตัวอย่างหรือประสบการณ์เพื่อการเรียนรู้งาน โดยมนุษย์มีส่วนร่วมเพียงการออกแบบระบบเท่านั้น หลังจากนั้นระบบจะสกัดสาระสำคัญจากตัวอย่างเหล่านี้เอง หลังจากการเรียนรู้เสร็จสิ้นด้วยตัวอย่างจำนวนหนึ่งอย่างเพียงพอ เครื่องหรือระบบที่เรียนรู้แล้วนี้สามารถนำไปใช้ในการประมวลผลของตัวอย่างใหม่ที่ไม่เคยพบมาก่อนได้อย่างมีประสิทธิภาพ

2.1.2 การเรียนรู้ของเครื่องและการเรียนรู้ของมนุษย์

การเรียนรู้ของมนุษย์ที่สามารถเป็นตัวอย่างที่ดีให้สามารถเข้าใจการเรียนรู้ของเครื่องได้ง่ายขึ้น เช่น ตอนวัยเด็กเราเรียนรู้การเขียนตัวอักษร ก.ไก่ ด้วยวิธีการคัดลายมือ ที่มีจุดประหรือจุดไข่ปลาให้เขียนตามได้ และเมื่อคล่องแล้วจะเริ่มคัดลายมือโดยไม่มีจุดไข่ปลาอีกต่อไป จนในที่สุดเราก็สามารถจดจำและทำการเขียนตัวอักษร ก.ไก่ ได้โดยไม่จำเป็นต้องทำการคัดลายมืออีกแล้วซึ่งเทียบกับการเรียนรู้ของเครื่องได้ดังนี้

- งาน (Task) คือ การเขียนตัวอักษร ก.ไก่ ได้
- ประสบการณ์ (Experience) คือ การฝึกคัดลายมือ
- ประสิทธิภาพ (Performance) คือ ความถูกต้องของลายมือที่เขียนอักษร ก.ไก่ หลังจากที่ได้ฝึกคัดลายมือแล้ว

นอกจากนี้ยังสามารถเปรียบเทียบโดยละเอียดให้ชัดยิ่งขึ้นได้ได้ดังนี้

- ชุดตัวอย่างฝึกฝน (Training Set) คือ ลายมือที่ถูกเขียนบนจุดไข่ปลา

- ปัญหาขนาดตัวอย่างน้อย (Small Sample Size (SSS) Problem) คือ คัดลายมือน้อยเกินไป
- การเรียนรู้แบบมีผู้สอน (Supervised Learning) คือ การมีจุดประหรือจุดไข่ปลาให้เขียนตามได้
- เป้าหมาย (Target) คือ จุดไข่ปลา
- ป้ายกำกับ (Label) คือ ตัวอักษร ก.ไก่
- ประสิทธิภาพ (Performance) วัดได้จาก ความผิดพลาด (Error) ที่เกิดขึ้นจากการคัดลายมือ
- ชุดประเมิน (Validation Set) คือ การคัดลายมือบนกระดาษที่ไม่มีจุดไข่ปลา
- ชุดทดสอบ (Test Set) คือ การเขียนอักษร ก.ไก่เพื่อใช้จริงในการประกอบกับตัวอักษรต่าง ๆ ให้เกิดคำ ซึ่งไม่ทราบล่วงหน้าว่าจะเขียนตัวอักษรใด

2.1.3 ลักษณะงานที่เหมาะสมกับการเรียนรู้ของเครื่อง

- งานที่มีความซับซ้อนมากหรือมีข้อมูลมากจนหาวิธีการวิเคราะห์หรือจัดการได้ยากด้วยมนุษย์
- งานที่มีปัจจัยและสภาพแวดล้อมมีการเปลี่ยนแปลงตลอดเวลา และต้องการเครื่องจักรที่สามารถปรับตัวให้เหมาะสมกับสภาพแวดล้อมนั้น ๆ ได้
- งานที่ไม่สามารถคาดการณ์ล่วงหน้าหรือไม่สามารถออกแบบวิธีการไว้ก่อนล่วงหน้าได้ เพราะต้องมีการตอบสนองที่แตกต่างกันไปตามสถานการณ์
- งานที่ไม่มีหรือไม่สามารถรู้รายละเอียดภายในระบบได้ ทราบเพียงแต่ข้อมูลนำเข้า (Input) และผลลัพธ์ (Output) เท่านั้น หรือบางครั้งเรียกว่าระบบ Black Box
- งานที่ต้องการวิเคราะห์ความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลจำนวนมาก ซึ่งไม่อาจทราบล่วงหน้าได้ถึงความสัมพันธ์นั้นว่าคือความสัมพันธ์ของสิ่งใดกับสิ่งใด
- งานที่ต้องอาศัยความรู้ที่เกิดขึ้นใหม่หรือเปลี่ยนแปลงอยู่ตลอดเวลา
- งานที่มีข้อมูลเข้าเดียวกัน แต่มีหลายรูปแบบหรือไม่คงที่
- งานที่ไม่สามารถแก้ปัญหาด้วยการค้นหาให้ครบถ้วนหรือครบทุกกรณีได้ (Brute-Force) หรือทำได้แต่ใช้เวลานานเกินกว่าจะยอมรับได้ในงานนั้น ๆ

2.1.4 ประเภทของการเรียนรู้ของเครื่อง

Machine Learning สามารถถูกแบ่งออกเป็นการเรียนรู้ได้ 2 แบบใหญ่ ๆ ได้แก่ การเรียนรู้แบบผู้สอน (Supervised Learning) และ การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

การเรียนรู้แบบมีผู้สอน (Supervised learning)

อัลกอริทึมจำเป็นต้องใช้ข้อมูลในส่วนการ train (training data) และส่วนที่รับกลับมาเพื่อปรับปรุง (feedback) จากมนุษย์เพื่อที่จะเรียนรู้ความสัมพันธ์ระหว่างข้อมูลที่ถูกป้อนเข้ามาสู่ข้อมูลที่ออกไป

กรณีของการเรียนรู้แบบมีผู้สอนเมื่อผลลัพธ์ของข้อมูลเป็นสิ่งที่รู้อยู่แล้ว อัลกอริทึมนี้ก็จะทำนายข้อมูลใหม่ได้ โดยมีประเภทของการเรียนรู้แบบมีผู้สอนอยู่ 2 ประเภท

- การแบ่งแยกประเภท (Classification)

  สมมติว่าเจ้าหน้าที่ต้องการทำนายเพศของลูกค้าซักคนหนึ่ง เจ้าหน้าที่จะเริ่มเก็บเกี่ยวข้อมูลตั้งแต่ ส่วนสูง น้ำหนัก อาชีพ เงินเดือน สินค้าที่ซื้อ และอื่น ๆ จากฐานข้อมูลลูกค้า การที่เจ้าหน้าที่รู้เพศของลูกค้าแต่ละคน มันสามารถตอบได้เพียงว่าลูกค้าผู้ชายหรือผู้หญิงเท่านั้น ผลลัพธ์จาก ตัวแบ่งแยกประเภท (classifier) นี้จะถูกกำหนดด้วยความน่าจะเป็นว่าเป็นผู้ชายหรือผู้หญิง (นั่นคือ การทำสัญลักษณ์ไว้ (label)) โดยขึ้นอยู่กับข้อมูลเป็นหลัก (นั่นคือ ลักษณะ (feature) ที่เราเก็บมา) เมื่อ model เรียนรู้ที่จะจดจำว่าเป็นผู้ชายหรือผู้หญิง คุณสามารถนำข้อมูลใหม่มาทำนายได้ เช่น เจ้าหน้าที่ได้รับข้อมูลใหม่จากลูกค้าแปลกหน้าและเจ้าหน้าที่ต้องการทราบว่าลูกค้าคนนั้นเป็นผู้ชายหรือผู้หญิง ถ้า classifier นี้ทำนายว่ามีโอกาสเป็นผู้ชาย 70% นั่นหมายความว่าอัลกอริทึมมีความแน่นอนอยู่ที่ 70% ที่ลูกค้าคนนั้นจะเป็นผู้ชายและ 30% ที่จะเป็นผู้หญิง

- การถดถอย (Regression)

  เมื่อผลลัพธ์มีค่าต่อเนื่องกันนั้นจะเป็นงานของ regression ที่จะมาช่วยแก้ปัญหานี้ เช่น นักวิเคราะห์ด้านการเงินคนหนึ่งอาจต้องการทำนายมูลค่าของหุ้นโดยดูจาก feature ต่าง ๆ เช่น ส่วนได้ส่วนเสีย (equity) สถานะของหุ้นก่อนหน้านี้ และดัชนีเศรษฐกิจมหภาค ระบบจะถูก train เพื่อประเมินราคาของหุ้นด้วยความผิดพลาดที่น้อยที่สุด

ตารางที่ 2.1 ข้อมูลอัลกอริทึมแบบต่าง ๆ

ชื่ออัลกอริทึม	คำอธิบาย	ประเภท
การถดถอยเชิงเส้น (Linear Regression)	หาความเกี่ยวข้องกันของแต่ละ Regression feature เพื่อหาผลลัพธ์เพื่อช่วย ทำนายมูลค่าในอนาคต	Regression
การถดถอยเชิงตรรกวิทยา (Logistic Regression)	เป็นส่วนเสริมของ Linear Regression แต่ใช้สำหรับ Classification โดยผลลัพธ์จะมีเพียง 2 คำตอบ เช่น ขาวหรือดำ	Classification
ต้นไม้ตัดสินใจ (Decision Tree)	สามารถแปลความได้หลายแบบไม่ว่าจะเป็น Regression หรือ Classification model ซึ่งแบ่งข้อมูลจาก feature เป็นแต่ละกิ่งของต้นไม้ ที่ตาของต้นไม้นั้น (decision nodes) (เช่น ถ้ามี feature เป็นสี ดังนั้นมีความเป็นไปได้ว่าจะมี feature สี เป็นกิ่งใหม่) จนกระทั่งการตัดสินครั้งสุดท้าย (final decision) ผลลัพธ์จะปรากฏขึ้น	Regression, Classification
Naïve Bayes	วิธีการของ Bayes เป็นวิธีการ Classification โดยมีการใช้ทฤษฎีบทของ Bayes ทฤษฎีบทดังกล่าว จะมีการปรับปรุงส่วนของความสำคัญ (prior knowledge) ของเหตุการณ์ด้วยความน่าจะเป็นที่เป็นอิสระต่อกันของแต่ละ feature ซึ่งสามารถมีผลต่อกันได้	Regression, Classification

ตารางที่ 2.1 ข้อมูลอัลกอริทึมแบบต่าง ๆ

ชื่ออัลกอริทึม	คำอธิบาย	ประเภท
ต้นไม้สุ่ม (Random Forest)	อัลกอริทึมนี้จะถูกสร้างขึ้นมาด้วย Decision tree เพื่อเพิ่มพูนความแม่นยำอย่างรุนแรง Random forest จะสร้าง Decision tree จำนวนมาก และใช้การโหวตโดยใช้เสียง ส่วนมาก (majority vote method) เพื่อตัดสินผลลัพธ์ที่จะปรากฏขึ้น สำหรับการ Classification ผลลัพธ์ของการพยากรณ์คือการผลโหวตที่มากที่สุดเพียงครั้งเดียว แต่ในการ Regression เป็นการเฉลี่ยคำ พยากรณ์ของต้นไม่ทุกต้นแล้วมา สรุปเป็นคำพยากรณ์สุดท้าย	Regression, Classification
Gradient-boosting trees	เป็นเทคนิคที่ทันสมัยที่สุดของการ Classification และ Regression มัน จะเน้นถึงเรื่องความผิดพลาดที่เกิดจากต้นไม้ก่อนหน้าและความพยายามที่จะแก้ไขมัน	Regression, Classification
Support Vector Machine (SVM)	ถูกใช้โดยทั่วไปสำหรับการ Classification อัลกอริทึมของ SVM จะหาเส้นแบ่งระหว่างสองประเภทที่ดีที่สุดออกมา โดยเส้นแบ่งนั้นจะเรียกว่า Hyperplane ซึ่งวิธีนี้เป็นวิธีที่ดีที่สุดสำหรับการแก้ปัญหากับ	Regression (ไม่ได้ใช้ทั่วไป), Classification

ตารางที่ 2.1 ข้อมูลอัลกอริทึมแบบต่าง ๆ

ชื่ออัลกอริทึม	คำอธิบาย	ประเภท
	โจทย์ที่ไม่ใช่เชิงเส้น (non-linear)	
Fast Large Margin	พัฒนาจาก fast margin learner โดย linear support vector learning แม้ว่าผลลัพธ์จะคล้ายกับผลลัพธ์ที่ได้จาก SVM แบบคลาสสิกหรือ logistic regression แต่ Fast Large Margin สามารถทำงานกับชุดข้อมูลที่มีตัวอย่างและคุณลักษณะหลายล้านรายการ	Regression, Classification
Deep Learning	วิธีการเรียนรู้แบบอัตโนมัติด้วยการเลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์ (Neurons) โดยนำระบบโครงข่ายประสาท (Neural Network) มาซ้อนกัน หลายชั้น (Layer) และทำการเรียนรู้ข้อมูลตัวอย่าง ซึ่งข้อมูลดังกล่าวจะถูกนำไปใช้ในการตรวจจับรูปแบบ (Pattern) หรือจัด หมวดหมู่ข้อมูล (Classify the Data)	Regression, Classification

<u>การเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning)</u>

อัลกอริทึมจะตรวจสอบเฉพาะข้อมูลที่ป้อนเข้ามาเท่านั้นโดยปราศจากการให้ผลลัพธ์ที่จะเกิดขึ้น (เช่น การสำรวจข้อมูลประชากรเพื่อหาแบบแผน (pattern) ของข้อมูลนั้น)

เราสามารถใช้การเรียนรู้แบบไม่มีผู้สอนเมื่อเราไม่ต้องการที่จะรู้ว่า machine แบ่งประเภทได้อย่างไร และเราต้องการอัลกอริทึมนั้นเพื่อหา pattern และแบ่งประเภทข้อมูล

2.1.5 คำศัพท์ที่ควรทราบเกี่ยวกับการเรียนรู้ของเครื่อง

- True Positives (TP) คือ ผลลัพธ์ที่แบบจำลองทำนายเชิงบวกได้อย่างถูกต้อง
- True Negatives (TN) คือ ผลลัพธ์ที่แบบจำลองทำนายเชิงลบได้อย่างถูกต้อง
- False Positives (FP) คือ ผลลัพธ์ที่แบบจำลองทำนายเชิงบวกไม่ถูกต้อง
- False Negatives (FN) คือ ผลลัพธ์ที่แบบจำลองทำนายเชิงลบไม่ถูกต้อง
- Accuracy (ความแม่น) คือ สัดส่วนของการคาดการณ์ว่าแบบจำลองที่ได้มานั้นถูกต้อง Accuracy สามารถคำนวณได้ดังนี้

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

หรือ

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precision (ความเที่ยง) คือ สัดส่วนของแบบจำลองทำนายเชิงบวกที่ถูกต้องจริง ๆ จากแบบจำลองทำนายเชิงบวกทั้งหมด สามารถนิยามได้ดังนี้

$$Precision = \frac{TP}{TP + FP}$$

- Recall (การระลึกได้) คือ สัดส่วนของแบบจำลองทำนายเชิงบวกที่ถูกต้องจริง ๆ ที่ได้รับการระบุอย่างถูกต้อง สามารถนิยามได้ดังนี้

$$Recall = \frac{TP}{TP + FN}$$

2.2 Particulate Matter 2.5 (PM 2.5)

ฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอน (PM 2.5) เป็นฝุ่นที่มีเส้นผ่านศูนย์กลางไม่เกิน 2.5 ไมครอน เกิดจากการเผาไหม้ทั้งจากยานพาหนะ การเผาวัสดุการเกษตร ไฟป่า และกระบวนการอุตสาหกรรม [8] สามารถเข้าไปถึงถุงลมในปอดได้ เป็นผลทำให้เกิดโรคในระบบทางเดินหายใจ และโรคปอดต่าง ๆ หากได้รับในปริมาณมากหรือเป็นเวลานานจะสะสมในเนื้อเยื่อปอด ทำให้การทำงานของปอดเสื่อมประสิทธิภาพลง ทำให้หลอดลมอักเสบ มีอาการหอบหืด PM 2.5 แบ่งเป็น 5 ระดับ คือ ตั้งแต่ 0 ถึง 201 ขึ้นไป ซึ่งแต่ละระดับจะใช้สีเป็นสัญลักษณ์เปรียบเทียบระดับของผลกระทบต่อสุขภาพอนามัย (ตารางที่ 2) โดยดัชนีคุณภาพอากาศ 100 จะมีค่าเทียบเท่ามาตรฐานคุณภาพอากาศในบรรยากาศโดยทั่วไป หากดัชนีคุณภาพอากาศมีค่าสูงเกินกว่า 100 แสดงว่าค่าความเข้มข้นของมลพิษทางอากาศมีค่าเกินมาตรฐานและคุณภาพอากาศในวันนั้นจะเริ่มมีผลกระทบต่อสุขภาพอนามัยของประชาชน

ตารางที่ 2.2 เกณฑ์ของดัชนีคุณภาพอากาศของประเทศไทย

AQI	ความหมาย	สีที่ใช้	คำอธิบาย
0 - 25	คุณภาพอากาศดีมาก	ฟ้า	คุณภาพอากาศดีมาก เหมาะสำหรับกิจกรรมกลางแจ้งและการท่องเที่ยว
26 - 50	คุณภาพอากาศดี	เขียว	คุณภาพอากาศดี สามารถทำกิจกรรมกลางแจ้งและการท่องเที่ยวได้ตามปกติ
51 - 100	ปานกลาง	เหลือง	ประชาชนทั่วไป : สามารถทำกิจกรรมกลางแจ้งได้ตามปกติ ผู้ที่ต้องดูแลสุขภาพเป็นพิเศษ : หากมีอาการเบื้องต้น เช่น ไอ หายใจลำบาก ระคายเคืองตา ควรลดระยะเวลาการทำกิจกรรมกลางแจ้ง
101 - 200	เริ่มมีผลกระทบต่อสุขภาพ	ส้ม	ประชาชนทั่วไป : ควรเฝ้าระวังสุขภาพ ถ้ามีอาการเบื้องต้น เช่น ไอ หายใจลำบาก ระคายเคืองตา ควรลดระยะเวลาการทำกิจกรรมกลางแจ้ง หรือใช้อุปกรณ์ป้องกันตนเองหากมีความจำเป็น ผู้ที่ต้องดูแลสุขภาพเป็นพิเศษ : ควรลดระยะเวลาการทำกิจกรรมกลางแจ้ง หรือใช้อุปกรณ์ป้องกันตนเองหากมีความจำเป็น ถ้ามีอาการทางสุขภาพ เช่น ไอ หายใจลำบาก ตาอักเสบ แน่นหน้าอก ปวดศีรษะ หัวใจเต้นไม่เป็นปกติ คลื่นไส้อ่อนเพลีย ควรปรึกษาแพทย์
201 ขึ้นไป	มีผลกระทบต่อสุขภาพ	แดง	ทุกคนควรหลีกเลี่ยงกิจกรรมกลางแจ้งหลีกเลี่ยงพื้นที่ที่มีมลพิษทางอากาศสูง หรือใช้อุปกรณ์ป้องกันตนเองหากมีความจำเป็น หากมีอาการทางสุขภาพควรปรึกษาแพทย์

PM 2.5 คือ ฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน อาจเทียบอย่างง่ายว่ามีขนาดประมาณ 1 ใน 25 ของเส้นผ่านศูนย์กลางของเส้นผมมนุษย์ เล็กขนาดที่ขนจมูกของมนุษย์นั้นไม่สามารถกรองได้ ทำให้ฝุ่นละอองชนิดนี้สามารถแพร่กระจายเข้าสู่ทางเดินหายใจ กระแสเลือด และแทรกซึมสู่กระบวนการทำงานในอวัยวะต่างๆ ของร่างกาย เพิ่มความเสี่ยงเป็นโรคเรื้อรัง

รศ.ดร.พิสุทธิ์ เพียรมนกุล อาจารย์ประจำภาควิชาวิศวกรรมสิ่งแวดล้อม คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และคณะกรรมการป้องกันและขจัดมลพิษทางน้ำเนื่องจากน้ำมัน (กปน.) ให้ความรู้ว่าปัญหา PM 2.5 ที่กำลังเผชิญกันอยู่นี้ เกิดจาก 4 สาเหตุ [10] คือ

1. แหล่งกำเนิด 4 แหล่ง ที่ก่อให้เกิด PM 2.5
   - การจราจร เป็นแหล่งกำเนิดใหญ่มีสัดส่วน 40% ของแหล่งกำเนิดทั้งหมด โดยเฉพาะเมืองใหญ่อย่างกรุงเทพฯ แล้ว เรียกว่าเป็นสัดส่วนหลัก เพราะกรุงเทพฯ รถติด และรถที่เคลื่อนที่ด้วยความเร็วต่ำหรือหยุดนิ่งอยู่กับที่ จะทำให้เกิด PM 2.5 เป็นจำนวนมาก
   - การเผาไหม้ในที่โล่ง เช่น เผาไร่ เผากระดาษ หรือแม้แต่การเผาไหม้ทางการเกษตรจากประเทศเพื่อนบ้าน ก่อให้มลพิษข้ามแดน จากปัจจัยนี้ทำให้เกิด PM 2.5 ไม่น้อยกว่า 30-40%
   - ภาคประชาชน กิจกรรมในชีวิตประจำวันบางประเภทที่เกิดเผาไหม้ ไม่ว่าจะเป็น สูบบุหรี่ ปิ้งย่าง สำหรับในเมืองใหญ่อย่างกรุงเทพฯ ที่มีประชากรกว่า 15 ล้านคน กิจกรรมเหล่านี้ก็มีส่วนทำให้เกิดมลภาวะเช่นกัน
   - ภาคอุตสาหกรรม ตามพระราชบัญญัติโรงงาน พ.ศ. 2535 และพระราชบัญญัติสิ่งแวดล้อม โรงงานต่าง ๆ จะต้องมีระบบบำบัดมลพิษทางอากาศ ก่อนจะปล่อยออกสู่ภายนอก องค์ประกอบที่จะทำ ให้เกิด PM 2.5 อย่างซัลเฟอร์ออกไซด์ ไนโตรเจนออกไซด์ ไฮโดรคาร์บอนที่เผาไม่สมบูรณ์ และความร้อน จึงเกิดขึ้นได้น้อย อย่างไรก็ตามโรงงานนั้นก็มีที่ปฏิบัติตามมาตรฐานและที่บกพร่อง ซึ่งเป็นเรื่องที่ภาครัฐต้องตรวจสอบต่อไป
2. สภาพอากาศ “สภาวการณ์ผันกลับของอุณหภูมิ (Temperature Inversion)” ที่มักเกิดขึ้นในฤดูหนาว ที่ชั้นความเย็นอยู่ภายใต้ถูกกักอยู่ภายใต้ชั้นความร้อน แปลงสภาพเป็นฝาชี กักกั้นให้ฝุ่นไม่สามารถลอยสูงได้ เป็นภาวะที่ประเทศเกาหลีและญี่ปุ่นประสบเช่นกัน
3. สภาพพื้นที่ ด้วยลักษณะทางภูมิศาสตร์เป็นแอ่งของกรุงเทพฯ ไม่เอื้อต่อการเจือจางมลพิษร่วมด้วยผังเมือง การวางตัว การก่อสร้างที่ทำให้การระบายอากาศไม่ดี
4. พื้นที่สีเขียว ตามองค์การอนามัยโลก (WHO) ระบุให้สัดส่วนพื้นที่สีเขียวต่อประชากร 1 คน คือ 9 ตร.ม. แต่สำหรับกรุงเทพมหานครแล้ว กลับมีสัดส่วนพื้นที่สีเขียวไม่ถึง 3 ตร.ม. ต่อประชากร 1 คน น้อยกว่ามาตรฐานของ WHO ถึง 3 เท่า

   องค์การอนามัยโลก (WHO) กำหนดให้ PM 2.5 จัดอยู่ในกลุ่มที่ 1 ของสารก่อมะเร็ง ตั้งแต่ปี พ.ศ. 2556 อีกทั้งยังเป็นสาเหตุให้ 1 ใน 8 ของประชากรโลกเสียชีวิตก่อนวัยอันควร

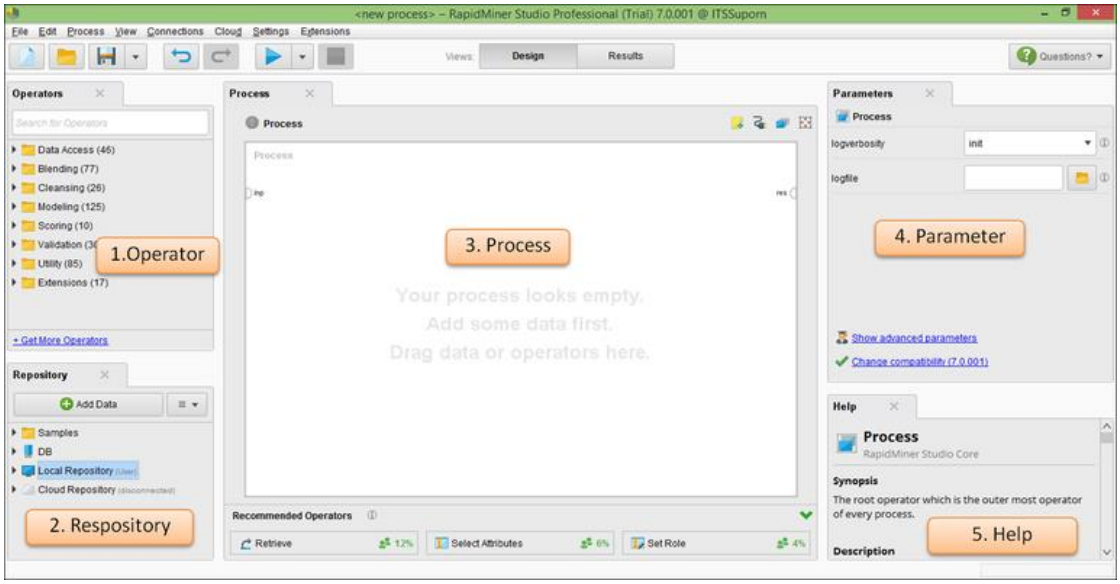
2.3 RapidMiner Studio

2.3.1 แนะนำ RapidMiner Studio

โลกในยุคปัจจุบันได้ก้าวเข้าไปสู่ยุคที่เรียกว่า “Big Data” หรือ “ข้อมูลอภิมหาศาล” เนื่องจากในแต่ละวันมีข้อมูลเกิดขึ้นมากมาย เช่น ข้อมูลสมาชิกของ Facebook ข้อมูลการซื้อสินค้าจากในซุปเปอร์มาร์เกตต่าง ๆ และเพื่อให้เกิดประโยชน์มากที่สุดบริษัทจึงจำเป็นต้องนำข้อมูลอภิมหาศาลเหล่านี้มาทำการวิเคราะห์ (analyze) ซึ่งเทคนิคหนึ่งที่ได้รับการนิยมอย่างสูงในปัจจุบัน คือ เทคนิค Data Mining ซึ่งเป็นเทคนิคที่ค้นหาความสัมพันธ์ในข้อมูล เช่น ถ้าลูกค้าซื้อเบียร์แล้วลูกค้าจะซื้อผ้าอ้อมร่วมไปด้วย หรือถ้าลูกค้ากด Like หน้า Facebook page ลูกค้าจะเห็นว่า Facebook มีระบบแนะนำ page อื่น ๆ ที่เกี่ยวข้องมาให้ด้วย หรือ การสร้างโมเดลเพื่อทำนายสิ่งที่จะเกิดขึ้นในอนาคต เช่น ทำนายยอดขายในไตรมาสถัดไป หรือ การทำนายว่าพนักงานคนไหนที่จะลาออกจากบริษัทในช่วง 3 เดือนข้างหน้า ตัวอย่างเหล่านี้ล้วนเป็นผลมาจากการวิเคราะห์ข้อมูลทางด้าน Data Mining

การวิเคราะห์ข้อมูลด้วย Data Mining นี้กำลังเป็นที่นิยมไปทั่วโลกด้วยแรงขับเคลื่อนอย่างหนึ่งคือ การมีซอฟต์แวร์ที่ช่วยให้ทำการวิเคราะห์ได้ง่ายขึ้น แต่ซอฟต์แวร์ส่วนใหญ่จะเป็นซอฟต์แวร์เชิงพาณิชย์ (commercial software) เช่น SAS Enterprise Miner หรือ IBM Intelligent Miner ทว่าการลงทุนซื้อซอฟต์แวร์เชิงธุรกิจเหล่านี้มาใช้งานอาจจะไม่คุ้มค่าในการลงทุนสำหรับผู้ประกอบการวิสาหกิจขนาดกลางและขนาดย่อม (SMEs) หรืออาจารย์ นักวิจัย และ นักศึกษาระดับปริญญาโทและเอก ในมหาวิทยาลัยต่าง ๆ ดังนั้นวิธีการหนึ่งที่จะทำให้เราสามารถวิเคราะห์ข้อมูลเหล่านี้ได้คือการใช้ open source software ที่สามารถดาวน์โหลดมาใช้งานได้โดยไม่เสียค่าใช้จ่าย เช่น ซอฟต์แวร์ RapidMiner Studio 8 ซึ่งเป็นซอฟต์แวร์ทาง Data Mining ที่ได้รับการโหวตว่ามีผู้ใช้งานมากที่สุดจากเว็บไซต์ KDnuggets.com เมื่อปี 2013 และตอนนี้ก็พัฒนามาถึง RapidMiner Studio 9.9

2.3.2 วิธีใช้งาน

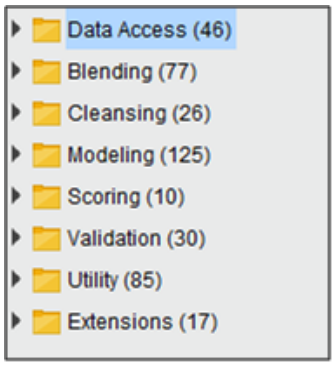


รูปที่ 2.1 หน้าต่างโปรแกรม RapidMiner Studio

จากรูปที่ 2.1 หน้า Design RapidMiner Studio ประกอบด้วย 5 ส่วนหลัก คือ

1. **Operators** เป็นส่วนที่ใช้เก็บตัวโอเปอร์เรเตอร์ที่ใช้ในการทำงานทั้งหมดซึ่งจัดเป็นกลุ่ม ๆ โดยกลุ่มที่ใช้งานคล้ายคลึงกันจะถูกจัดอยู่ในกลุ่มเดียวกัน และสรุปได้ว่ามี 8 กลุ่มโอเปอร์เรเตอร์คือ

1. Data Access
2. Blending
3. Cleansing
4. Modeling
5. Scoring
6. Validation
7. Utility
8. Extensions

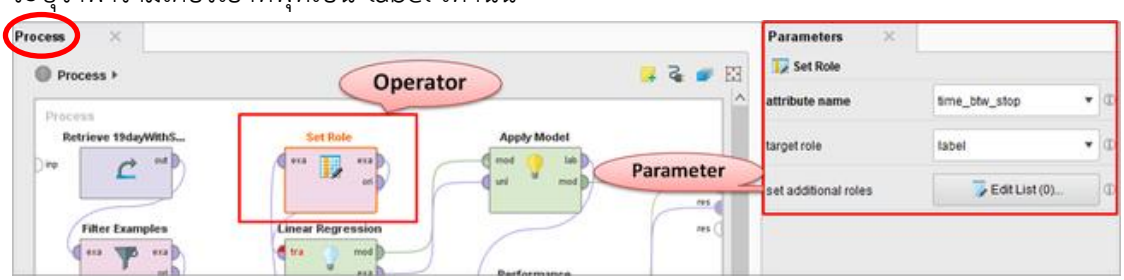


รูปที่ 2.2 หน้าต่าง Operators

2. **Repository** ส่วนนี้เป็นส่วนจัดการไฟล์ RapidMiner Studio จะจัดการข้อมูลจาก 3 แหล่งคือ DB (ดาต้าเบส) Local (ในเครื่องคอมพิวเตอร์ที่ใช้อยู่) และ Cloud Repository (ในคลาวด์) โดยเก็บไฟล์ Data Set และ Process ต่าง ๆ แยกเก็บคนละโฟลเดอร์กัน

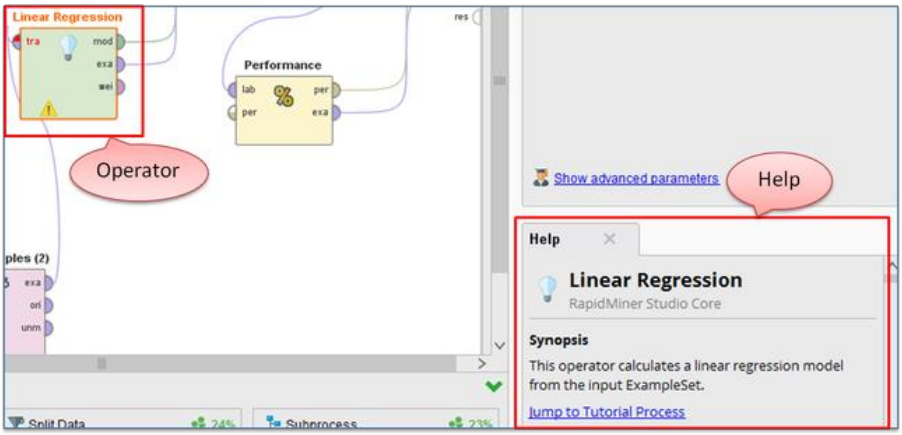
3. **Process** เป็นหน้าหลักในการทำงานในการสร้างโปรเซสสำหรับทำ Machine Learning ของซอฟต์แวร์นี้ โดยจะนำโอเปอร์เรเตอร์มาประกอบเพื่อสร้างโปรเซสขึ้นมาตามวัตถุประสงค์ของโจทย์ที่ตั้งไว้

4. **Parameters** เป็นส่วนที่กำหนดพารามิเตอร์ที่เป็นรายละเอียดของ โอเปอร์เรเตอร์ ที่เลือกใช้งาน เช่น โอเปอร์เรเตอร์ Set Role เป็น Operator ที่มีพารามิเตอร์ที่เกี่ยวข้องสองพารามิเตอร์คือ แอททริบิวต์เนม ซึ่งเป็นพารามิเตอร์ที่ใช้กำหนดเอาท์พุทของโปรเซสว่าจะเลือกใหน ซึ่งจากตัวอย่าง เลือก time_btw_stop เป็นเอาท์พุทของโปรเซส และ พารามิเตอร์ target role เพื่อระบุว่าพารามิเตอร์เอาท์พุทเป็น label เท่านั้น



รูปที่ 2.3 หน้าต่าง Process และ Parameters ของหัวข้อที่ 3 และ 4

5. **Help** เป็นส่วนช่วยเหลือซึ่งจะแสดงรายละเอียดของตัวโอเปอร์เรเตอร์ที่เลือกใช้งานอยู่ ส่วนช่วยเหลือของ RapidMiner Studio 7 จะบอกเพียงหน้าที่และรายละเอียดคร่าว ๆ ของโอเปอร์เรเตอร์หากต้องการดูรายละเอียดมากกว่านี้ต้องไปที่ Jum to Tutorai Process ซึ่งจะ Link ไปยังเวปไซต์ที่มีรายละเอียดของเกี่ยวกับ Operator ที่ใช้อยู่ เช่น โอเปอร์เรเตอร์ชื่อ Linear Regression ในหน้า Help ก็จะบอกว่าเป็น โอเปอร์เรเตอร์ ที่ใช้คำนวณข้อมูลจากจาก data set โดยใช้วิธี Linear Regression



รูปที่ 2.4 หน้าต่าง Help ของหัวข้อที่ 5

ส่วนของ Auto Model ของ RapidMiner Studio สามารถดูได้ที่บทที่ 3

บทที่ 3
วิธีการวิจัย

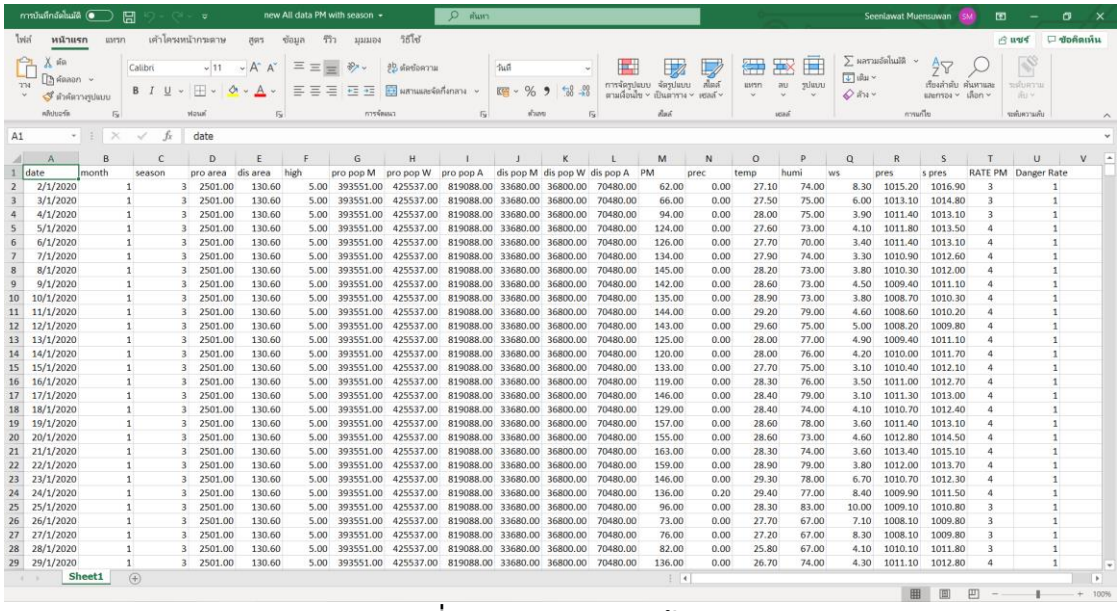
บทนี้จะกล่าวถึงวิธีการวิจัยการประมาณค่าความหนาแน่นของ PM 2.5 ด้วยการเรียนรู้ของเครื่อง โดยการใช้โปรแกรม RapidMiner Studio ในการวิเคราะห์ข้อมูลของ PM 2.5 ด้วยการดูความสัมพันธ์ระหว่าง PM 2.5 กับปัจจัยอื่น ๆ

3.1 วิเคราะห์ข้อมูล PM 2.5 และการคาดคะเน

จากบทที่ 2 ผู้วิจัยสามารถคาดคะเนได้ว่ามีปัจจัยที่ส่งผลต่อปริมาณ PM 2.5 คือ ฤดูกาล พื้นที่ต่าง ๆ ความสูงจากระดับน้ำทะเล จำนวนประชากร ปริมาณฝน อุณหภูมิ ความชื้น ความเร็วลม ความกดอากาศ และความกดอากาศจากระดับน้ำทะเล

3.2 เก็บรวบรวมข้อมูล

ผู้วิจัยดำเนินการหา และเก็บรวบรวมข้อมูลจากแหล่งข้อมูลต่าง ๆ เช่น ข้อมูลเกี่ยวกับอุตุนิยมวิทยา เช่น ปริมาณฝน อุณหภูมิ ความชื้น ผู้วิจัยก็ไปขอข้อมูลจากกรมอุตุนิยมวิทยา และเว็บไซต์ Thai Meteorlogical Department หรือข้อมูล PM 2.5 ผู้วิจัยก็ขอจากกรมควบคุมมลพิษ และเว็บไซต์ Air Quality Historical Data Platform หลังจากนั้นผู้วิจัยก็นำข้อมูลที่ได้มาจัดเรียงและการทำความสะอาดข้อมูล โดยทางผู้วิจัยกำหนดให้ค่า PM 2.5 สีฟ้ากับสีเขียวมีค่า Dangerous Rate เป็น 0 ซึ่งหมายความว่าไม่ต้องใส่หน้ากากอนามัย และกำหนดให้ค่า PM 2.5 สีเหลือง สีส้ม และสีแดง มีค่า Dangerous Rate เป็น 1 ซึ่งหมายความว่าต้องใส่หน้ากากอนามัย เพื่อให้ง่ายต่อการวิเคราะห์ด้วยโปรแกรม นำข้อมูลที่มีมาลงในโปรแกรม Microsoft Excel ตัวอย่างข้อมูลแสดงดังรูปที่ 3.1



date	month	season	pro area	dis area	high	pro pop M	pro pop W	pro pop A	dis pop M	dis pop W	dis pop A	PM	prec	temp	humi	ws	pres	s pres	RATE PM	Danger Rate
2/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	62.00	0.00	27.10	74.00	8.30	1015.20	1016.90	3	1
3/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	66.00	0.00	27.50	75.00	6.00	1013.10	1014.80	3	1
4/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	94.00	0.00	28.00	75.00	3.90	1011.40	1013.10	3	1
5/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	124.00	0.00	27.60	73.00	4.10	1011.80	1013.50	4	1
6/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	126.00	0.00	27.70	70.00	3.40	1011.40	1013.10	4	1
7/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	134.00	0.00	27.90	74.00	3.30	1010.90	1012.60	4	1
8/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	145.00	0.00	28.20	73.00	3.80	1010.30	1012.00	4	1
9/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	142.00	0.00	28.60	73.00	4.50	1009.40	1011.10	4	1
10/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	135.00	0.00	28.90	73.00	3.80	1008.70	1010.30	4	1
11/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	144.00	0.00	29.20	79.00	4.60	1008.60	1010.20	4	1
12/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	143.00	0.00	29.60	75.00	5.00	1008.20	1009.80	4	1
13/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	125.00	0.00	28.00	77.00	4.90	1009.40	1011.10	4	1
14/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	120.00	0.00	28.00	76.00	4.20	1010.00	1011.70	4	1
15/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	133.00	0.00	27.70	75.00	3.10	1010.40	1012.10	4	1
16/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	119.00	0.00	28.30	76.00	3.50	1011.00	1012.70	4	1
17/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	146.00	0.00	28.40	79.00	3.10	1011.30	1013.00	4	1
18/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	129.00	0.00	28.40	74.00	4.10	1010.70	1012.40	4	1
19/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	157.00	0.00	28.60	78.00	3.60	1011.40	1013.10	4	1
20/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	155.00	0.00	28.60	73.00	4.60	1012.80	1014.50	4	1
21/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	163.00	0.00	28.30	74.00	3.60	1013.40	1015.10	4	1
22/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	159.00	0.00	28.90	79.00	3.80	1012.00	1013.70	4	1
23/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	146.00	0.00	29.30	78.00	6.70	1010.70	1012.30	4	1
24/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	136.00	0.20	29.40	77.00	8.40	1009.90	1011.50	4	1
25/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	96.00	0.00	28.30	83.00	10.00	1009.10	1010.80	3	1
26/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	73.00	0.00	27.70	67.00	7.10	1008.10	1009.80	3	1
27/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	76.00	0.00	27.20	67.00	8.30	1008.10	1009.80	3	1
28/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	82.00	0.00	25.80	67.00	4.10	1010.10	1011.80	3	1
29/1/2020	1	3	2501.00	130.60	5.00	393551.00	425537.00	819088.00	33680.00	36800.00	70480.00	136.00	0.00	26.70	74.00	4.30	1011.10	1012.80	4	1

รูปที่ 3.1 Excel ตารางข้อมูล

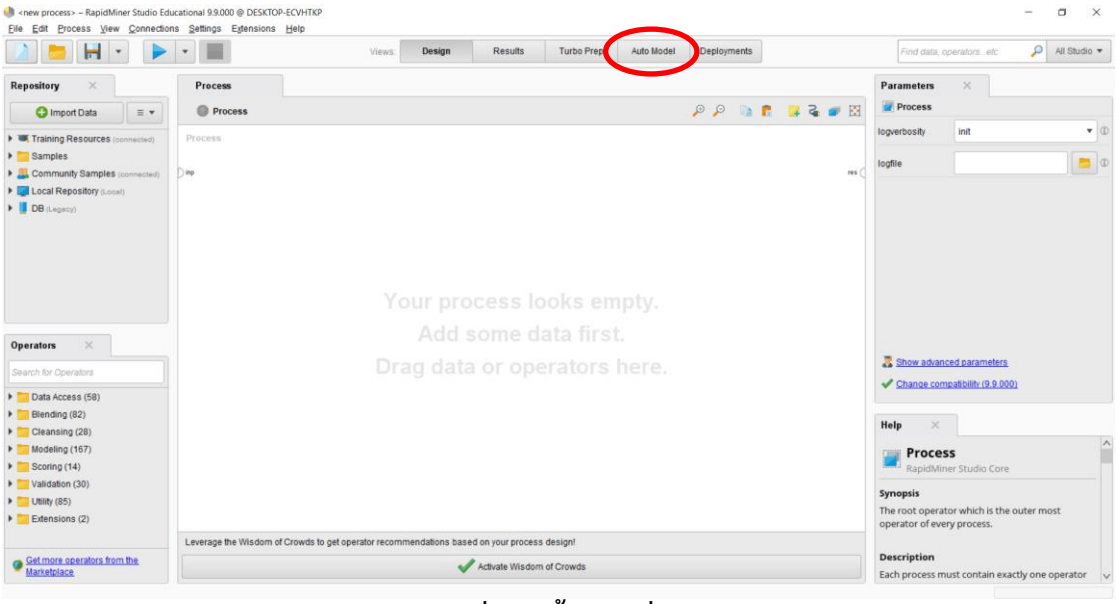
โดย date คือ วันที่, month คือ เดือน, season คือ ฤดูกาล 1 เป็นฤดูหนาว 2 เป็นฤดูร้อน 3 เป็นฤดูฝน, pro area คือ พื้นที่จังหวัด, dis area คือ พื้นที่อำเภอ, high คือความสูงจากระดับน้ำทะเล, pro pop M คือ ประชากรชายในจังหวัด, pro pop W คือ ประชากรหญิงในจังหวัด, pro pop A คือ ประชากรรวมในจังหวัด, dis pop M คือ ประชากรชายในอำเภอ, dis pop W คือ ประชากรหญิงในอำเภอ, dis pop A คือ ประชากรรวมในอำเภอ, PM คือค่า PM 2.5, prec คือปริมาณหยาดน้ำฟ้า, temp คือ อุณหภูมิ, humi คือความชื้น, ws คือ ความเร็วลม, pres คือความกดอากาศ, s pres คือความกดอากาศจากระดับน้ำทะเล, RATE PM คือ ค่าของสี PM ซึ่งสีฟ้า เป็น 1 สีเขียวเป็น 2 สีเหลืองเป็น 3 สีส้มเป็น 4 สีแดงเป็น 5 และ Danger Rate คือ ค่าความอันตราย ซึ่ง 0 คือไม่ต้องใส่หน้ากากอนามัย และ 1 คือ ต้องใส่หน้ากากอนามัย

3.3 นำข้อมูลมาวิเคราะห์ด้วยโปรแกรม

ในช่วงแรกทางผู้วิจัยคาดว่าความสัมพันธ์จะสามารถหาได้เป็นโมเดลถดถอย หรือ Regression แต่เมื่อนำข้อมูลมาลองวิเคราะด้วยโปรแกรม SPSSv22.0 ก็ไม่สามารถหาค่าทางสถิติที่เหมาะสมในการตอบโจทย์ที่ต้องการได้ จึงเปลี่ยนโปรแกรมที่ใช้วิเคราะห์มาเป็น RapidMiner Studio โดยใช้ Machine Learning ในการวิเคราะห์ และทางผู้วิจัยได้ใช้ฟังก์ชัน Auto Model เพื่อหาโมเดลที่ดีที่สุด โดยฟังก์ชันนี้เป็นฟังก์ชันที่โปรแกรมจะดำเนินการเลือกโมเดลให้แก่ผู้ใช้งานเองจากการวิเคราะห์ข้อมูลนำเข้าที่กำหนดให้

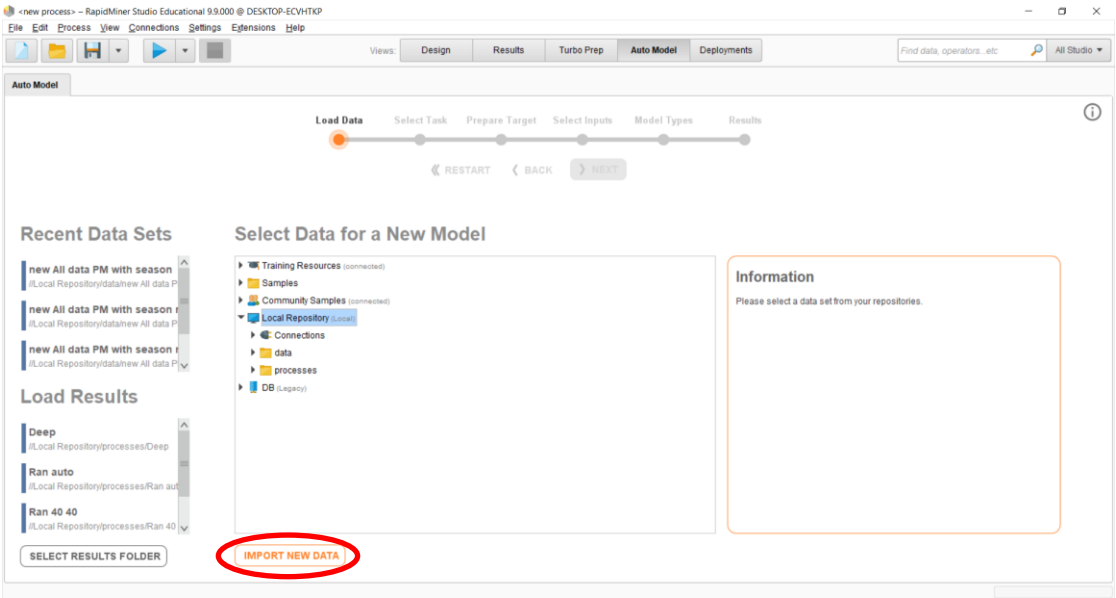
ขั้นตอนในการดำเนินการวิเคราะห์เพื่อหาโมเดลด้วยฟังก์ชัน Auto Model มีรายละเอียดดังนี้

วิธีการใช้ฟังก์ชัน Auto Model ของ RapidMiner Studio



รูปที่ 3.2 ขั้นตอนที่ 1

- จากรูปที่ 3.2 กดที่ปุ่ม Auto Model จะได้จอภาพปรากฏดังรูปที่ 3.3

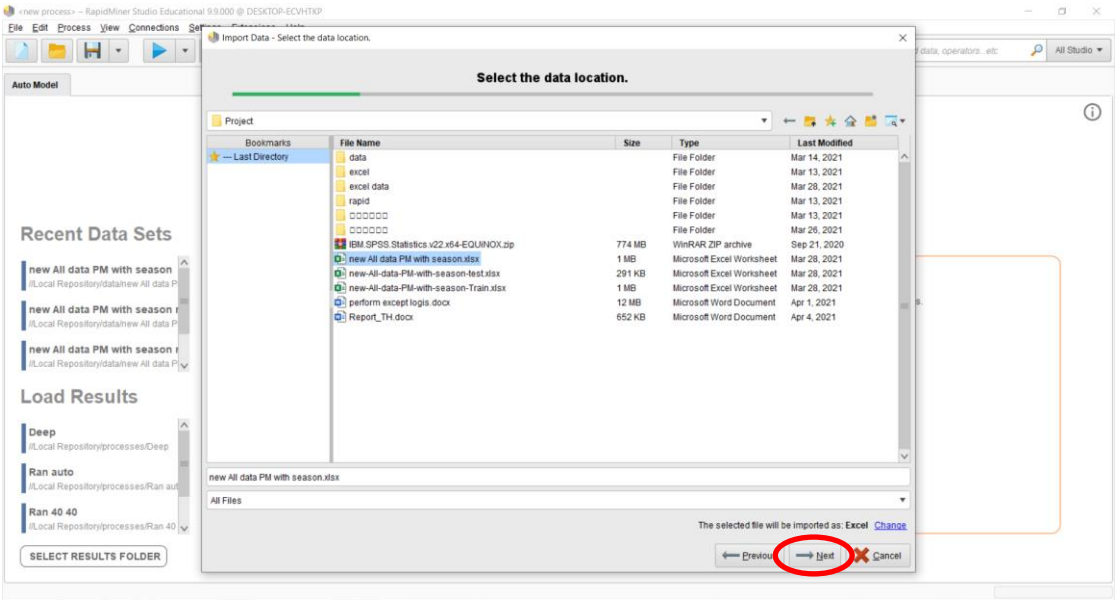


รูปที่ 3.3 ขั้นตอนที่ 2

- จากรูปที่ 3.3 กดที่ปุ่ม IMPORT NEW DATA จะได้จอภาพปรากฏดังรูปที่ 3.4

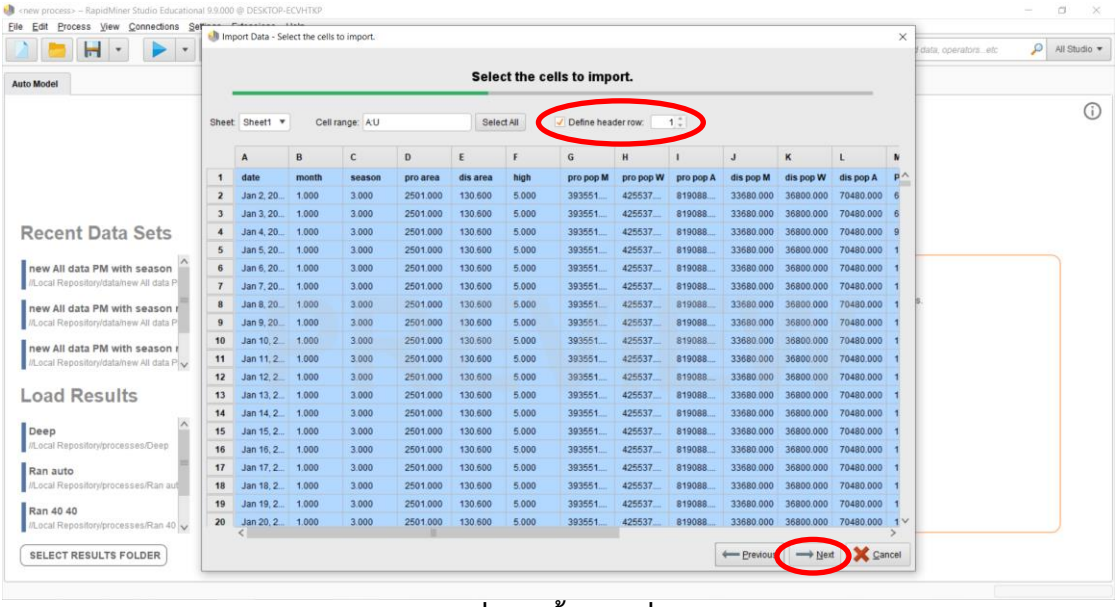
รูปที่ 3.4 ขั้นตอนที่ 3

- จากรูปที่ 3.4 เลือกแหล่งไฟล์ของข้อมูลที่เราจะนำมาใช้ หลังกดปุ่มแหล่งไฟล์ที่เลือกแล้วจะได้จอภาพปรากฏดังรูปที่ 3.5



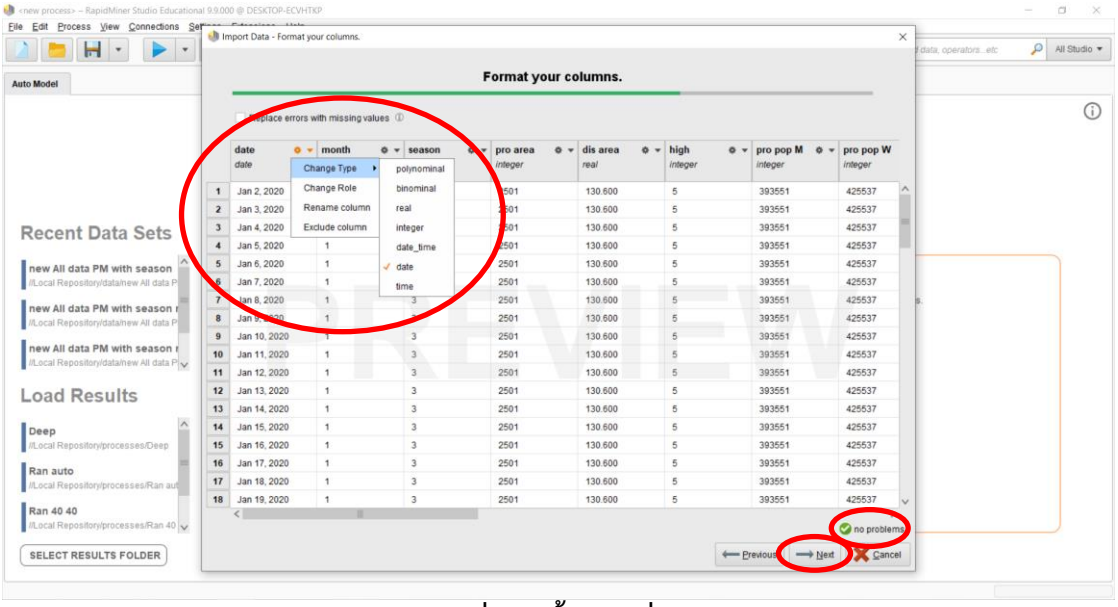
รูปที่ 3.5 ขั้นตอนที่ 4

- จากรูปที่ 3.5 เมื่อเลือกไฟล์ได้แล้ว ให้กดปุ่ม Next จะได้จอภาพปรากฏดังรูปที่ 3.6



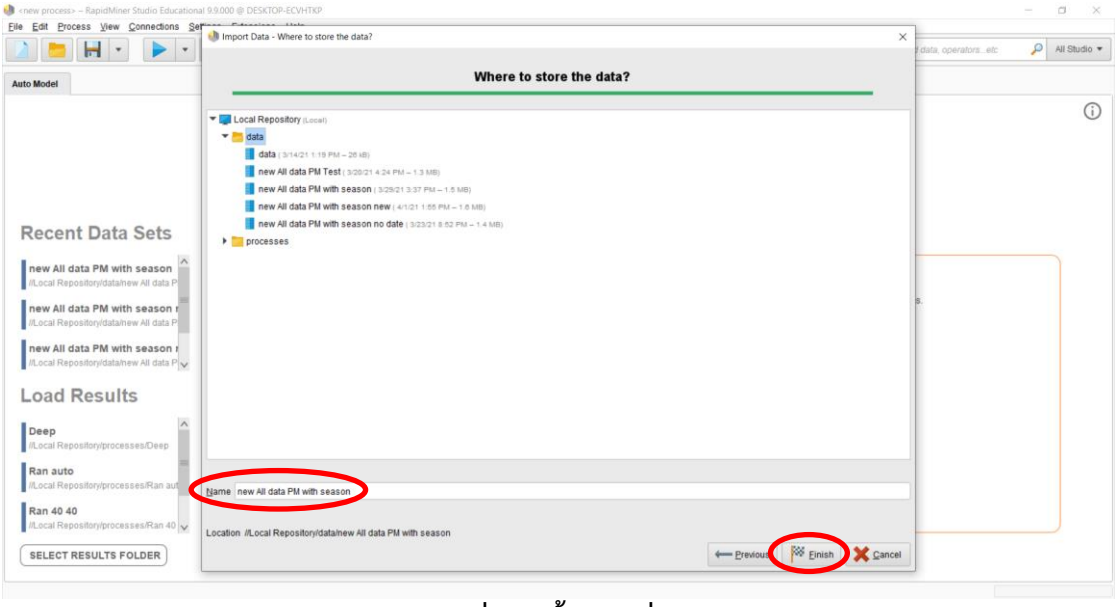
รูปที่ 3.6 ขั้นตอนที่ 5

- จากรูปที่ 3.6 ตรวจสอบว่าข้อมูลที่นำเข้ามา มีปัญหาหรือไม่ และนำเข้าข้อมูลครบหรือไม่ รวมทั้งกำหนดว่าหัวข้อคือแถวไหนที่ช่อง Define header row หลังจากนั้นกดปุ่ม Next จะได้จอภาพปรากฏดังรูปที่ 3.7



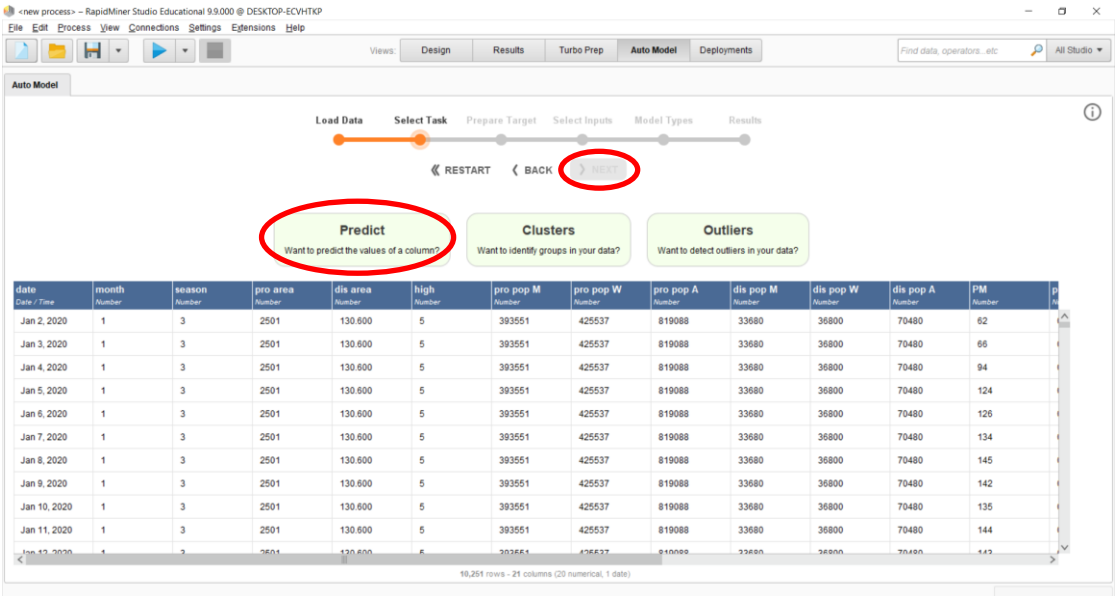
รูปที่ 3.7 ขั้นตอนที่ 6

- จากรูปที่ 3.7 ตรวจสอบดูว่าประเภทของข้อมูลถูกต้องหรือไม่ ถ้าไม่ถูกต้องเราสามารถเปลี่ยนประเภทได้ มีปัญหาหรือไม่ ถ้ามีปัญหา จะขึ้นว่า have problems ที่มุมขวาล่าง ถ้าไม่มีปัญหา จะขึ้นว่า no problems ที่มุมขวาล่าง หลังจากที่ตรวจสอบแล้วให้กดปุ่ม Next จะได้จอภาพปรากฏดังรูปที่ 3.8



รูปที่ 3.8 ขั้นตอนที่ 7

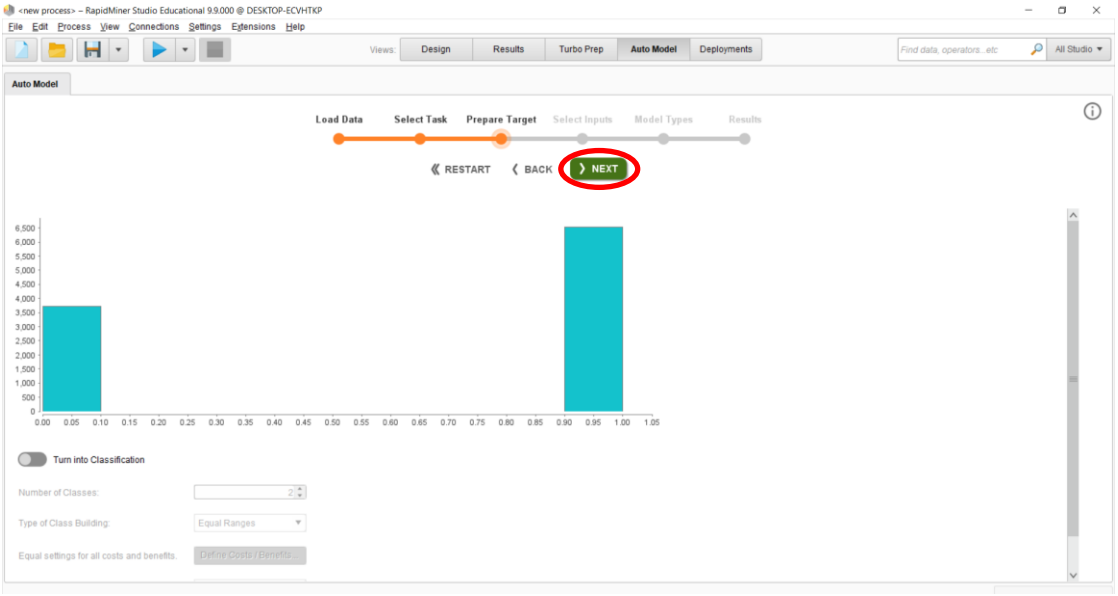
- จากรูปที่ 3.8 เลือกตำแหน่งที่จะเซฟไฟล์ แล้วตั้งชื่อไฟล์ใหม่ หลังจากนั้นกดปุ่ม Finish จะได้จอภาพปรากฏดังรูปที่ 3.9



date	month	season	pro area	dis area	high	pro pop M	pro pop W	pro pop A	dis pop M	dis pop W	dis pop A	PM
Jan 2, 2020	1	3	2501	130.600	5	393551	425537	819088	33680	36800	70480	62
Jan 3, 2020	1	3	2501	130.600	5	393551	425537	819088	33680	36800	70480	66
Jan 4, 2020	1	3	2501	130.600	5	393551	425537	819088	33680	36800	70480	94
Jan 5, 2020	1	3	2501	130.600	5	393551	425537	819088	33680	36800	70480	124
Jan 6, 2020	1	3	2501	130.600	5	393551	425537	819088	33680	36800	70480	126
Jan 7, 2020	1	3	2501	130.600	5	393551	425537	819088	33680	36800	70480	134
Jan 8, 2020	1	3	2501	130.600	5	393551	425537	819088	33680	36800	70480	145
Jan 9, 2020	1	3	2501	130.600	5	393551	425537	819088	33680	36800	70480	142
Jan 10, 2020	1	3	2501	130.600	5	393551	425537	819088	33680	36800	70480	135
Jan 11, 2020	1	3	2501	130.600	5	393551	425537	819088	33680	36800	70480	144

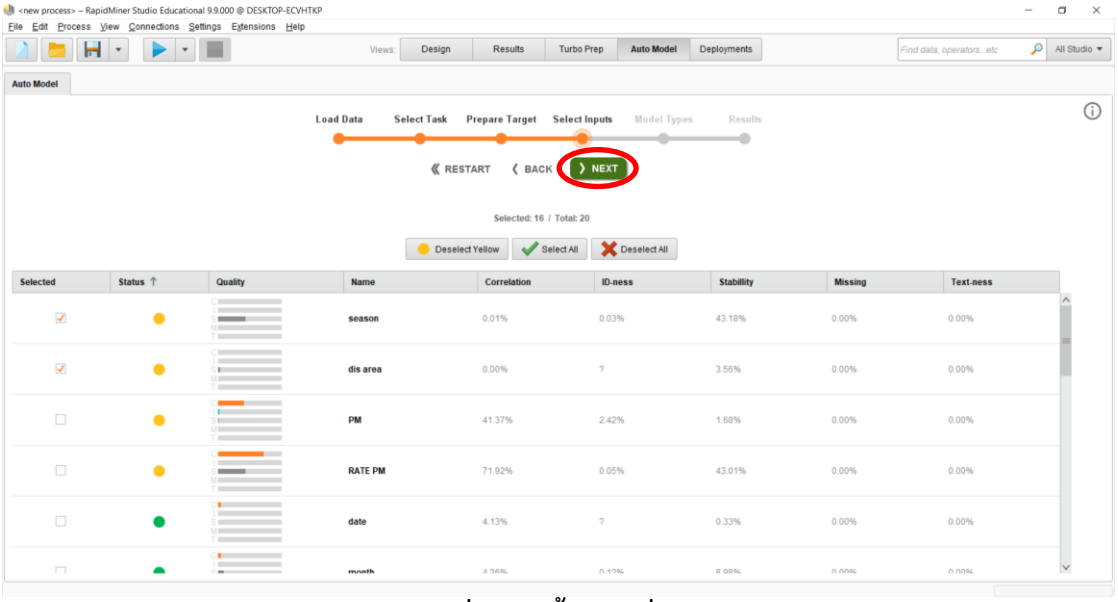
รูปที่ 3.9 ขั้นตอนที่ 8

- จากรูปที่ 3.9 เลือกว่าเราจะทำอะไรกับข้อมูล หลังจากนั้นกดปุ่ม Next ซึ่งในที่นี้จะเลือก Predict เพื่อจะทำนายว่าค่าความอันตรายของ PM 2.5 เป็นอย่างไรบ้างจากความสัมพันธ์ของปัจจัยต่าง ๆ กับ PM 2.5 จะได้จอภาพปรากฏดังรูปที่ 3.10



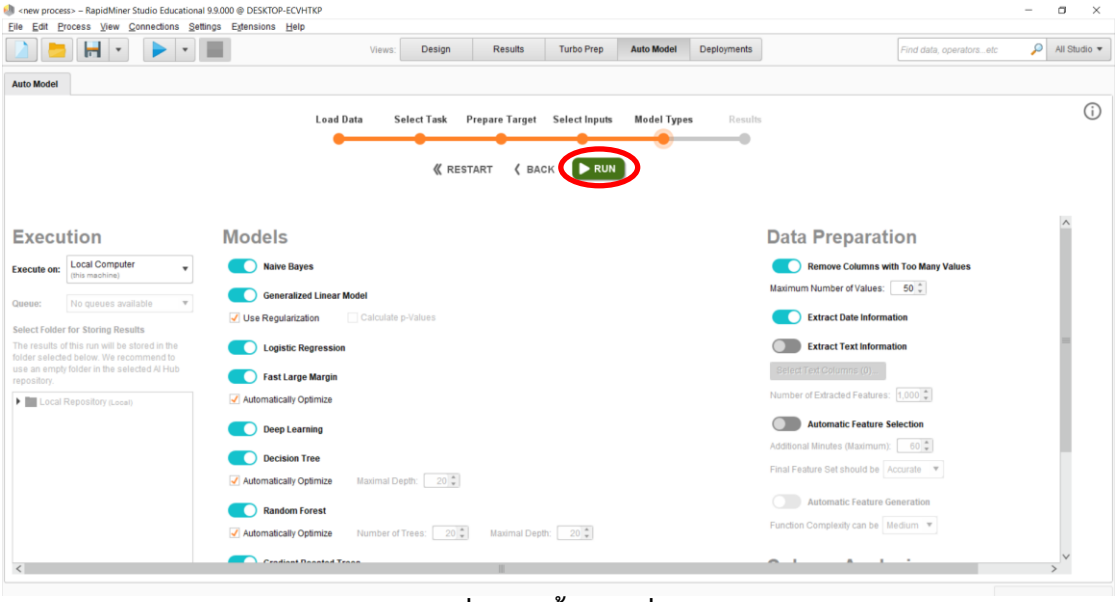
รูปที่ 3.10 ขั้นตอนที่ 9

- จากรูปที่ 3.10 กดปุ่ม Next อีกครั้ง จะได้จอภาพปรากฏดังรูปที่ 3.11



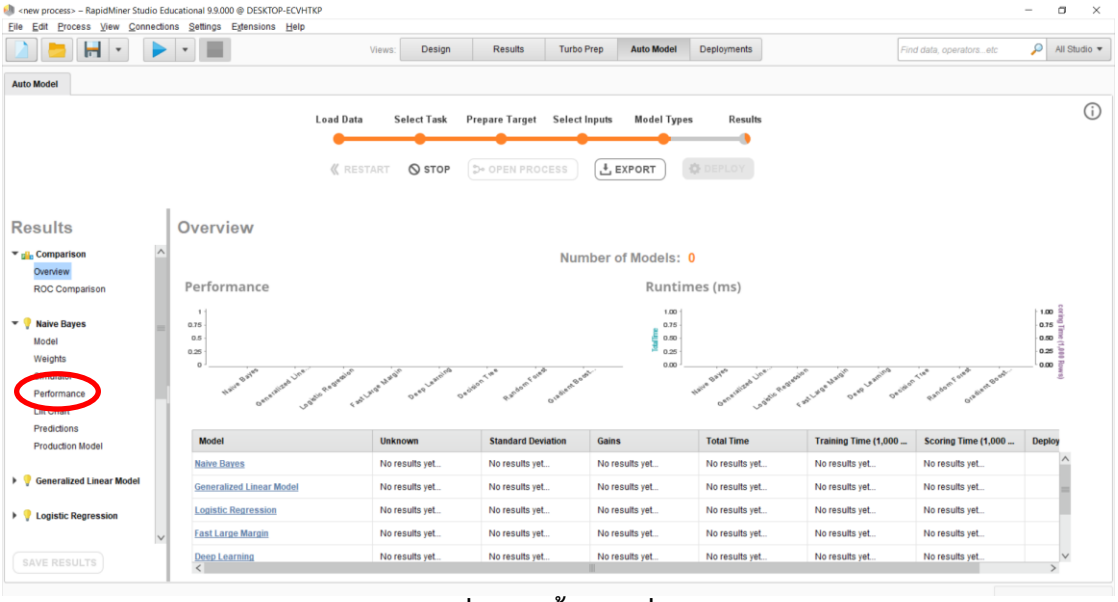
รูปที่ 3.11 ขั้นตอนที่ 10

- จากรูปที่ 3.11 เลือกข้อมูลที่จะนำมาวิเคราะห์ โดยคลิ๊กเครื่องหมาย ✓ ในช่องที่ต้องการเลือก หลังจากนั้นกดปุ่ม Next จะได้จอภาพปรากฏดังรูปที่ 3.12



รูปที่ 3.12 ขั้นตอนที่ 11

- จากรูปที่ 3.12 เลือกโมเดลที่ต้องการทดสอบ ว่าต้องการจะทดสอบด้วยโมเดลใดบ้าง หลังจากนั้นกดปุ่ม Run จะได้จอภาพปรากฏดังรูปที่ 3.13



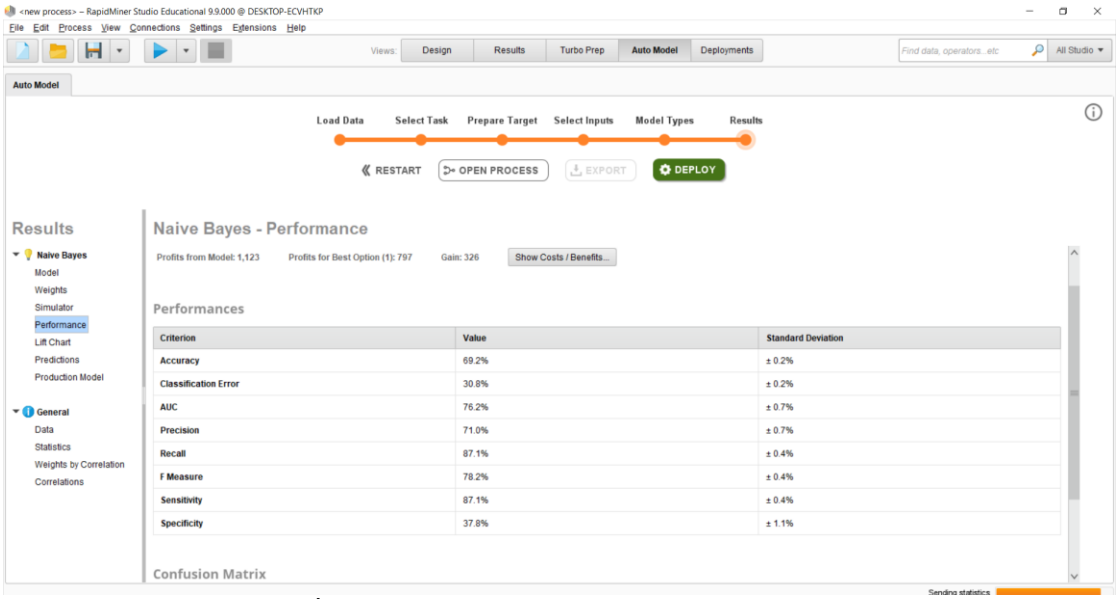
รูปที่ 3.13 ขั้นตอนที่ 12

- จากรูปที่ 3.13 รอให้โปรแกรมรัน แล้วพิจารณาว่าโปรแกรมไหนดีที่สุดจาก Performance ของโมเดลต่าง ๆ โดยกดที่ปุ่ม Performance ของแต่ละโมเดล จะได้ผลลัพธ์ดังหัวข้อที่ 3.4

3.4 การเลือกโมเดลที่ที่ดีสุด

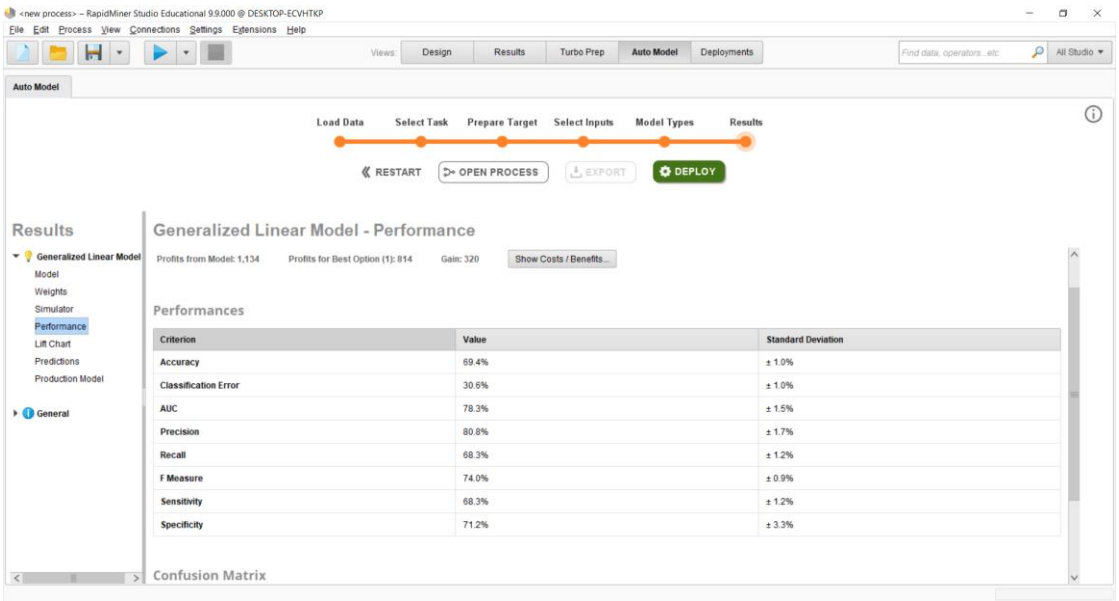
ผู้วิจัยได้ใช้ Auto Model ของ RapidMiner Studio เพื่อหาโมเดลที่มีประสิทธิภาพมากที่สุด จากค่าของ Performance ของแต่ละโมเดล

- Naïve Bayes



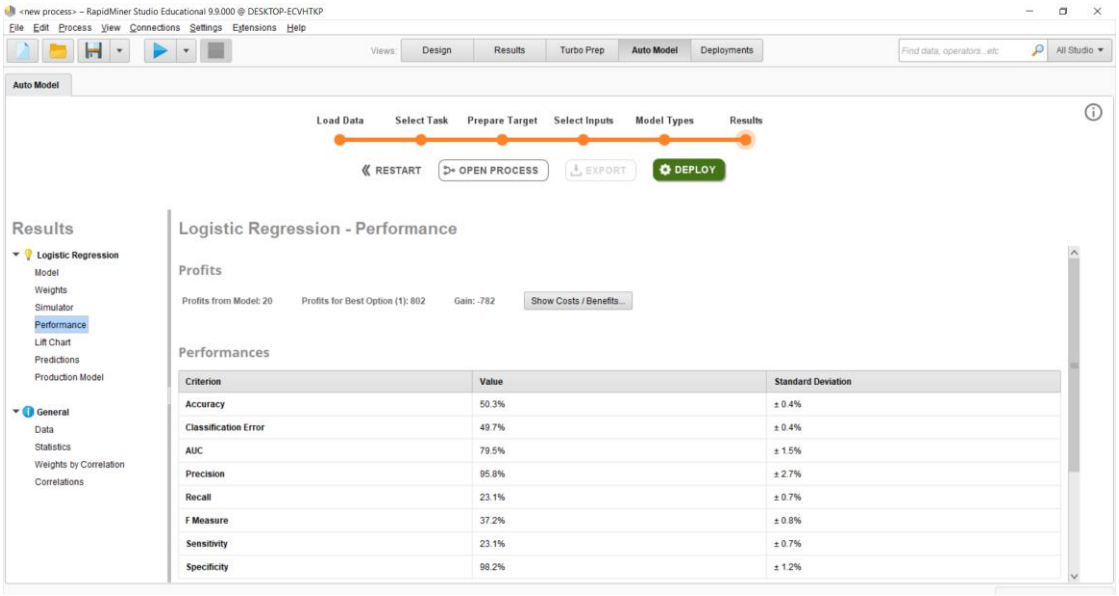
รูปที่ 3.14 Performance ของ Naïve Bayes

- Generalized Lenear Model



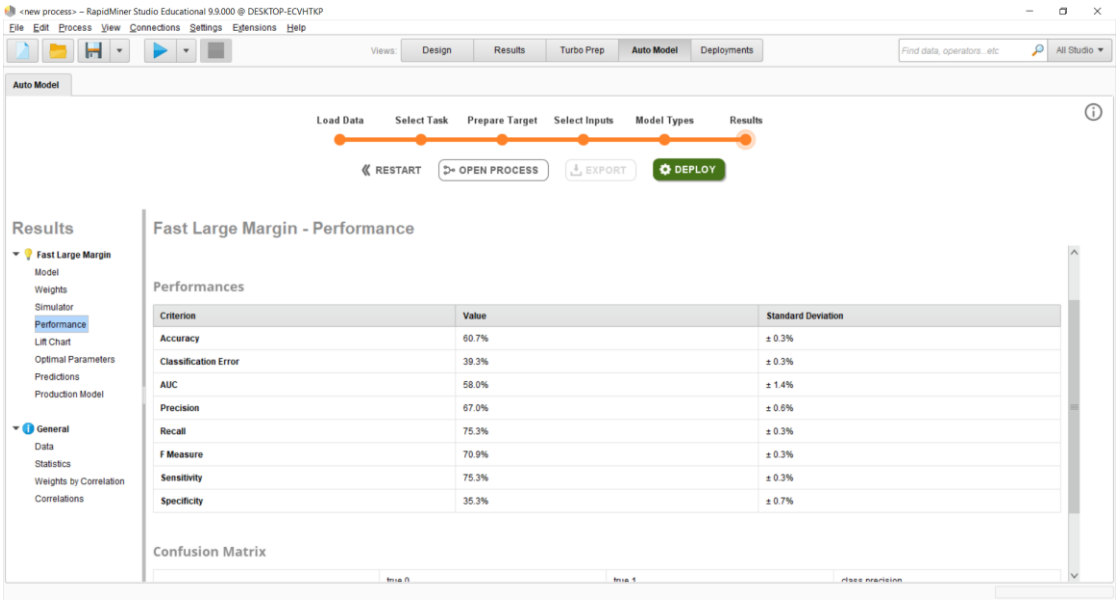
รูปที่ 3.15 Performance ของ Generalized Lenear Model

- Logistic Regression



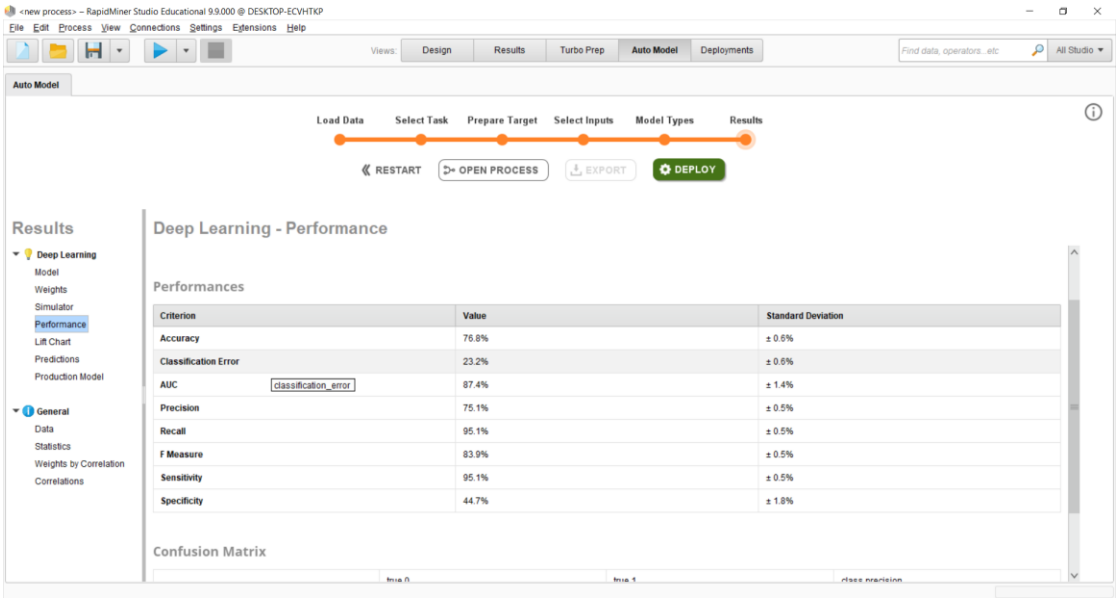
รูปที่ 3.16 Performance ของ Logistic Regression

- Fast Large Margin



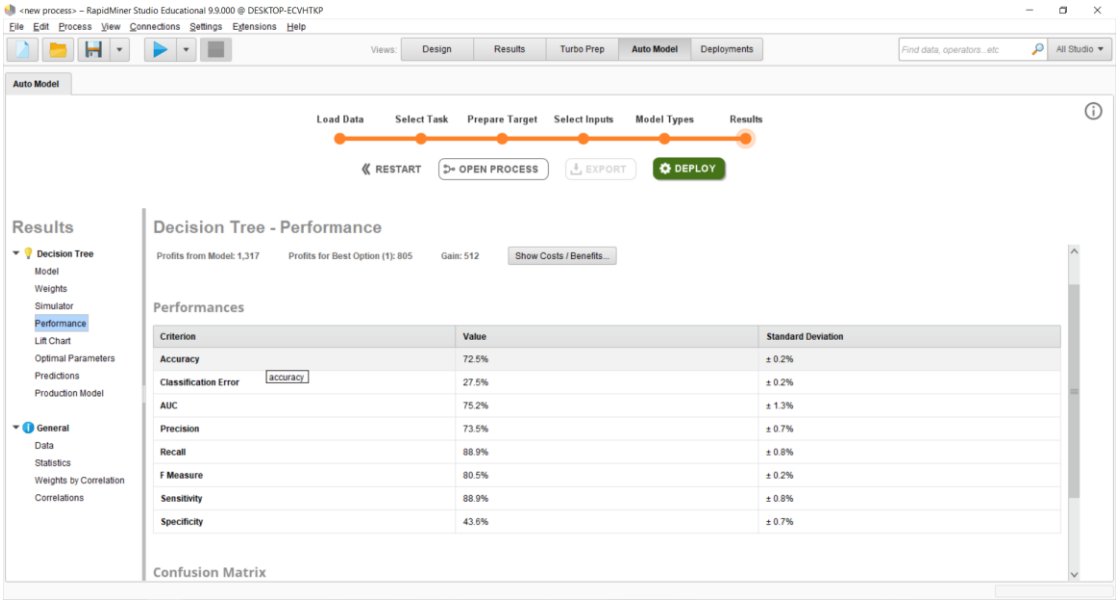
รูปที่ 3.17 Performance ของ Fast Large Margin

- Deep Learning



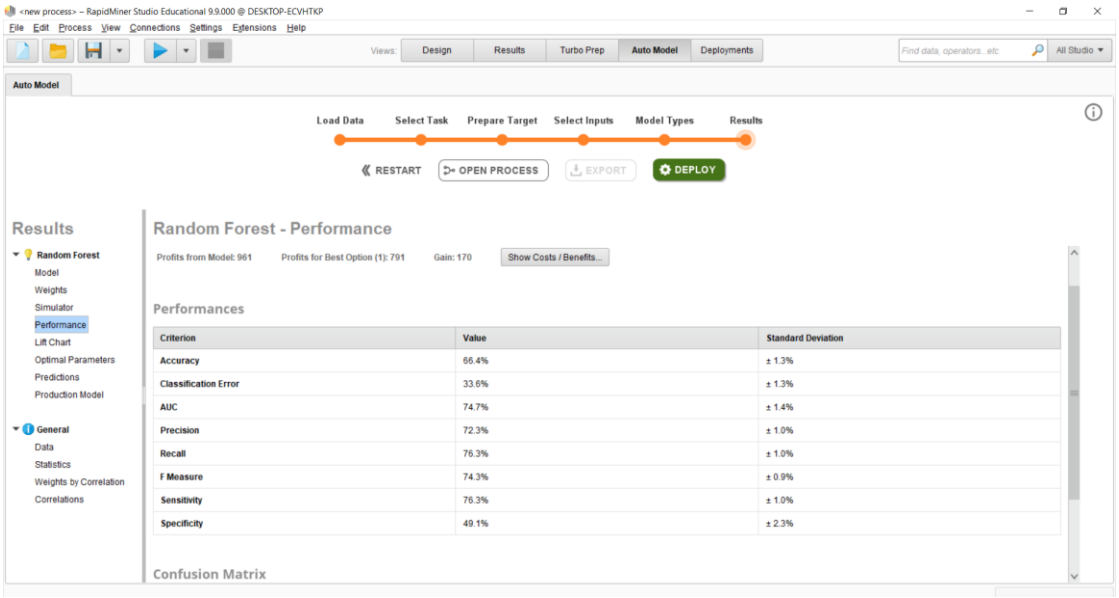
รูปที่ 3.18 Performance ของ Deep Learning

- Decision Tree



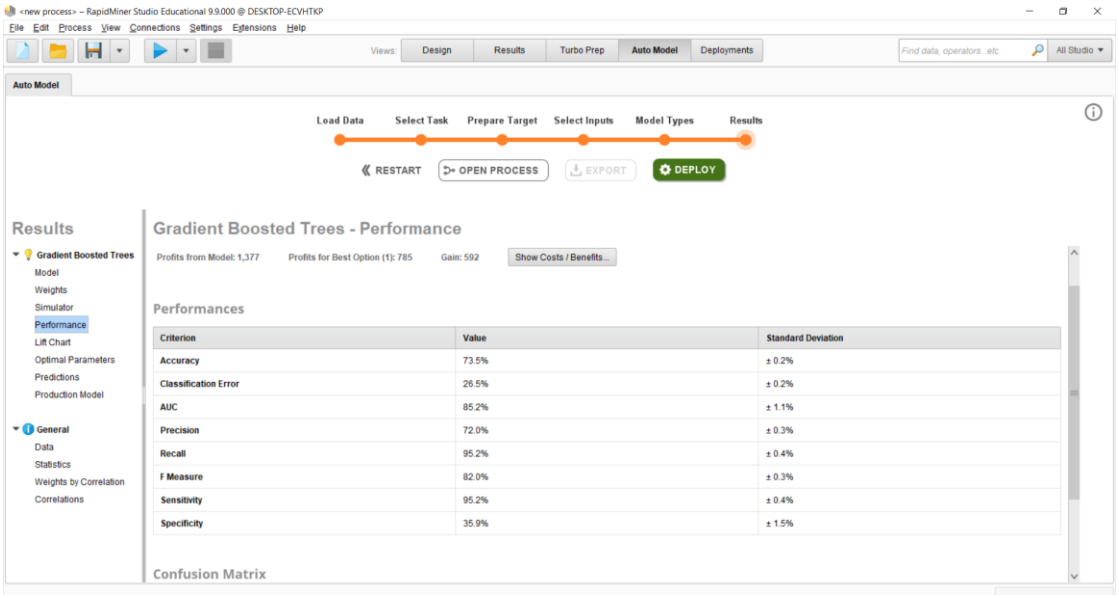
รูปที่ 3.19 Performance ของ Decision Tree

- Random Forest



รูปที่ 3.20 Performance ของ Random Forest

- Gradient Boosted Trees



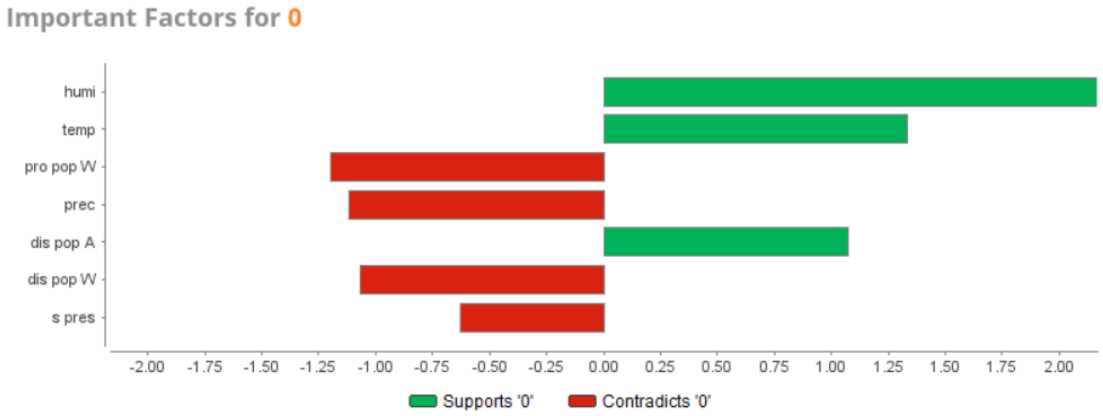
รูปที่ 3.21 Performance ของ Gradient Boosted Trees

จากการพิจารณา Performance ของทุกโมเดลที่ได้กล่าวข้างต้น สามารถสรุปได้ว่า โมเดล Deep Learning เป็นโมเดลที่ดีที่สุด จากการที่มีค่า Recall และ Accuracy สูงที่สุด คือ 95.2% และ 76.8% ตามลำดับ หลังจากนั้น ผู้วิจัยได้นำโมเดล Deep Learning มาวิเคราะห์ความสัมพันธ์ของค่า PM 2.5 กับปัจจัยต่าง ๆ เพื่อดูว่าปัจจัยใดส่งผลอย่างไรกับ PM 2.5 บ้าง

บทที่ 4
ผลการวิจัย

บทนี้จะกล่าวถึง ผลของการประมาณค่าความหนาแน่นของ PM 2.5 ด้วยการเรียนรู้ของเครื่อง ด้วยโมเดล Deep Learning ซึ่งได้ผลการวิจัยว่า

4.1 กรณีที่ประชาชนไม่ต้องใส่หน้ากากอนามัย (0)



รูปที่ 4.1 ปัจจัยที่สำคัญต่อกรณีที่ประชาชนไม่ต้องใส่หน้ากากอนามัย (0)

จากรูปที่ 4.1 ให้ผลลัพธ์ว่า

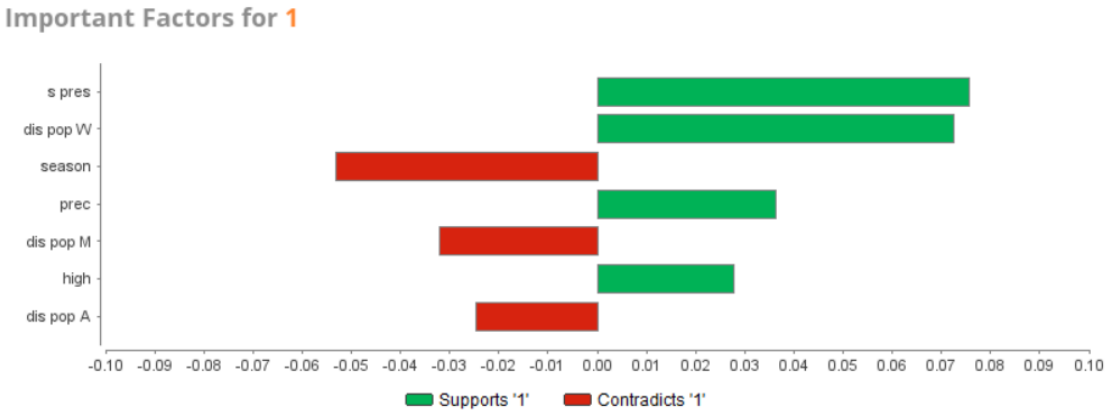
<u>ปัจจัยที่ส่งผลในทางสนับสนุนให้ประชาชนไม่ต้องใส่หน้ากากอนามัย คือ</u>

- ความชื้น ทำให้ปริมาณ PM 2.5 ในอากาศลดลง
- อุณหภูมิ ทำให้ปริมาณ PM 2.5 ในอากาศลดลง
- จำนวนประชากรรวมของอำเภอ ทำให้ปริมาณ PM 2.5 ในอากาศลดลง

<u>ปัจจัยที่ส่งผลในทางตรงข้ามให้ประชาชนไม่ต้องใส่หน้ากากอนามัย คือ</u>

- จำนวนประชากรหญิงของจังหวัด ทำให้ปริมาณ PM 2.5 ในอากาศมากขึ้น
- ปริมาณหยาดน้ำฟ้า ทำให้ปริมาณ PM 2.5 ในอากาศมากขึ้น
- จำนวนประชากรหญิงของอำเภอ ทำให้ปริมาณ PM 2.5 ในอากาศมากขึ้น
- ความกดอากาศที่ระดับน้ำทะเล ทำให้ปริมาณ PM 2.5 ในอากาศมากขึ้น

4.2 กรณีที่ประชาชนควรใส่หน้ากากอนามัย (1)



รูปที่ 4.2 ปัจจัยที่สำคัญต่อกรณีที่ประชาชนควรใส่หน้ากากอนามัย (1)

จากรูปที่ 4.2 ให้ผลลัพธ์ว่า

<u>ปัจจัยที่ส่งผลในทางสนับสนุนให้ประชาชนควรใส่หน้ากากอนามัย คือ</u>

- ความกดอากาศที่ระดับน้ำทะเล ทำให้ปริมาณ PM 2.5 ในอากาศมากขึ้น
- จำนวนประชากรหญิงของอำเภอ ทำให้ปริมาณ PM 2.5 ในอากาศมากขึ้น
- ปริมาณหยาดน้ำฟ้า ทำให้ปริมาณ PM 2.5 ในอากาศมากขึ้น
- ความสูงจากระดับน้ำทะเล ทำให้ปริมาณ PM 2.5 ในอากาศมากขึ้น

<u>ปัจจัยที่ส่งผลในทางตรงข้ามให้ประชาชนควรใส่หน้ากากอนามัย คือ</u>

- ฤดูกาล ทำให้ปริมาณ PM 2.5 ในอากาศลดลง
- จำนวนประชากรชายของอำเภอ ทำให้ปริมาณ PM 2.5 ในอากาศลดลง
- จำนวนประชากรรวมของอำเภอ ทำให้ปริมาณ PM 2.5 ในอากาศลดลง

บทที่ 5
ข้อสรุปและข้อเสนอแนะ

ในบทนี้จะกล่าวถึง สรุปผลจากข้อมูลผลการวิจัยในบทที่ 4 ข้อสรุปการประมาณค่าความหนาแน่นของ PM 2.5 ด้วยการเรียนรู้ของเครื่อง รวมถึงข้อเสนอแนะอื่นๆ ที่ได้จากการวิจัย

5.1 ข้อสรุป

จากหัวข้อ 2.2 Particulate Matter 2.5 ข้อมูลจาก รศ.ดร.พิสุทธิ์ เพียรมนกุล รวมกับข้อมูลกรณีที่ประชาชนไม่ต้องใส่หน้ากากอนามัย (0) และ กรณีที่ประชาชนควรใส่หน้ากากอนามัย (1) พบว่าปัจจัยที่ส่งผลให้ประชาชนไม่ต้องใส่หน้ากากอนามัยหรือค่าความอันตรายของ PM 2.5 ต่ำ คือ

- ความชื้น ทำให้ปริมาณ PM 2.5 ในอากาศลดลง เนื่องจากความชื้นไปจับ PM 2.5 ให้ลงสู่พื้นดิน
- อุณหภูมิ ทำให้ปริมาณ PM 2.5 ในอากาศลดลง เนื่องจากเมื่ออุณหภูมิสูงขึ้น ทำให้อากาศลอยตัวสูงขึ้น ทำให้ PM 2.5 ถูกพัดออกไปง่ายขึ้น
- จำนวนประชากรชายของอำเภอ ทำให้ปริมาณ PM 2.5 ในอากาศลดลง เนื่องจากกิจกรรมในชีวิตประจำวันของประชากรผู้ชาย เช่น การประกอบอาชีพ ชาวสวน ชาวไร่ ซึ่งเป็นการเพิ่มพื้นที่สีเขียว
- จำนวนประชากรรวมของอำเภอ ทำให้ปริมาณ PM 2.5 ในอากาศลดลง เนื่องจากกิจกรรมในชีวิตประจำวันของประชากรผู้ชาย ที่ทำให้ PM 2.5 ลดลงกว่ากิจกรรมในชีวิตประจำวันของประชากรผู้หญิง
- ฤดูกาล ทำให้ปริมาณ PM 2.5 ในอากาศเพิ่มขึ้น เนื่องจาก“สภาวการณ์ผันกลับของอุณหภูมิ (Temperature Inversion)” ที่มักเกิดขึ้นในฤดูหนาว ที่ชั้นความเย็นอยู่ภายใต้ถูกกักอยู่ภายใต้ชั้นความร้อน แปลงสภาพเป็นฝาชี กักกั้นให้ฝุ่นไม่สามารถลอยสูงได้ เป็นภาวะที่ประเทศเกาหลีและญี่ปุ่นประสบเช่นกัน

นอกจากนี้ยังพบว่า ปัจจัยที่ส่งผลให้ประชาชนควรใส่หน้ากากอนามัยหรือค่าความอันตรายของ PM 2.5 สูง คือ

- จำนวนประชากรหญิงของอำเภอ ทำให้ปริมาณ PM 2.5 ในอากาศมากขึ้น เนื่องจากกิจกรรมในชีวิตประจำวันของประชากรผู้หญิง เช่น การประกอบอาชีพแม่ค้า แม่ครัว

- จำนวนประชากรหญิงของจังหวัด ทำให้ปริมาณ PM 2.5 ในอากาศมากขึ้น เนื่องจากกิจกรรมในชีวิตประจำวันของประชากรผู้หญิง เช่น การประกอบอาชีพแม่ค้า แม่ครัว
- ปริมาณหยาดน้ำฟ้า ทำให้ปริมาณ PM 2.5 ในอากาศมากขึ้น เนื่องจากตัวเครื่องวัดแยกระหว่างฝุ่นละออง กับละอองฝนไม่ออก
- ความกดอากาศที่ระดับน้ำทะเล ทำให้ปริมาณ PM 2.5 ในอากาศมากขึ้น เนื่องจากความกดอากาศทำให้ PM 2.5 ลอยอยู่ในระดับต่ำ และไม่สามารถถูกพัดไปไหนได้
- ความสูงจากระดับน้ำทะเล ทำให้ปริมาณ PM 2.5 ในอากาศมากขึ้น เนื่องจากความสูงยิ่งมากอากาศจะเบาบางลง ทำให้ PM 2.5 ลอยอยู่ได้นานขึ้น

ปัจจัยอื่นอาจส่งผลต่อค่าความอันตรายของ PM 2.5 เช่นกัน แต่อาจมีผลน้อยถึงน้อยมาก ทำให้โปรแกรมไม่แสดงออกมา

5.2 ข้อเสนอแนะ

ผู้วิจัยไม่ได้นำข้อมูลของยานพาหนะ และการก่อสร้างมารวมในข้อมูลด้วย ทำให้ยังขาดปัจจัยในด้านนี้อยู่ เนื่องจากข้อมูลในที่ใช้เยอะแล้ว และโครงงานนี้ก็เน้นไปทางเรื่องอุตุนิยมวิทยา อีกทั้งเวลาไม่พอไปค้นหาและเก็บรวบรวมข้อมูลของยานภาหนะ และการก่อสร้างมาเพิ่ม

รายการอ้างอิง

1. ABB. Deep Learning คืออะไร? [ออนไลน์]. 2563. แหล่งที่มา : https://new.abb.com/news/detail/58004/deep-learning [01 เมษายน 2564]
2. BBC NEWS | ไทย. ฝุ่น PM 2.5 กลับมาวิกฤต รัฐบาลทำอะไรไปแล้วบ้างหลังประกาศเป็น “วาระแห่งชาติ” [ออนไลน์]. 2563. แหล่งที่มา : https://www.bbc.com/thai/thailand-51030236 [01 พฤศจิกายน 2563]
3. BBC NEWS | ไทย. ฝุ่น : เหตุใดสถานการณ์ฝุ่นละอองขนาดเล็ก จึงพุ่งสูงขึ้นมาอีกครั้ง [ออนไลน์]. 2561. แหล่งที่มา : https://www.bbc.com/thai/thailand-46643980 [01 พฤศจิกายน 2563]
4. Living sneak peek. PM 2.5 คืออะไร? [ออนไลน์]. 2563. แหล่งที่มา : https://www.livingsneakpeek.com/pm-2-5-คืออะไร/ [01 พฤศจิกายน 2563]
5. Machine Learning Crash Course. Classification: Precision and Recall [ออนไลน์]. 2561. แหล่งที่มา : https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall [01 เมษายน 2564]
6. rapidminer. Fast Large Margin [ออนไลน์]. 2564. แหล่งที่มา : https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/support_vector_machines/fast_large_margin.html [01 เมษายน 2564]
7. THE STANDARD. PM2.5 ฝุ่นละอองขนาดเล็กในอากาศ กับวิกฤตสุขภาพที่คนไทยจะต้องแลก [ออนไลน์]. 2561. แหล่งที่มา : https://thestandard.co/pm-2-5-environmental-nano-pollutants/ [01 เมษายน 2564]
8. กรมควบคุมมลพิษ. ข้อมูลดัชนีคุณภาพอากาศ [ออนไลน์]. แหล่งที่มา : http://air4thai.pcd.go.th/webV2/aqi_info.php [01 พฤศจิกายน 2563]
9. กรมควบคุมมลพิษ. มาตรการป้องกันและแก้ไขปัญหาฝุ่นละออง PM 2.5 [ออนไลน์]. 2562. แหล่งที่มา : https://www.pdmo.go.th/pdmomedia/documents/2020/Jan/มาตรการป้องกันและแก้ไขปัญหาฝุ่นละออง%20PM%202.pdf [01 พฤศจิกายน 2563]
10. ข่าวสด. สาเหตุที่แท้ของ PM 2.5 เรียนรู้ เข้าใจ ป้องกัน ไม่ตื่นตระหนก [ออนไลน์]. 2562. แหล่งที่มา : https://www.khaosod.co.th/pr-news/news_2171737 [01 เมษายน 2564]

11. ประชาชาติธุรกิจ. ศูนย์วิจัยกสิกรไทยชี้ “ฝุ่น PM2.5 ที่รุนแรงขึ้น” ส่อกระทบเศรษฐกิจสูงถึง 6 พันล้านบาท [ออนไลน์]. 2563. แหล่งที่มา : https://www.prachachat.net/finance/news-413478 [01 พฤศจิกายน 2563]
12. รศ. ดร.ปริญญา สงวนสัตย์, Artificial Intelligence with Machone Learning AI สร้างได้ด้วยแมชชีนเลิร์นนิ่ง Python Edition, (นนทบุรี : บริษัท ไอดีซี พรีเมียร์ จำกัด, 2562), หน้า 3-12.
13. สมาคมโปรแกรมเมอร์ไทย. อะไรคือ การเรียนรู้ของเครื่อง (Machine Learning)? (ฉบับมือใหม่) [ออนไลน์]. 2561. แหล่งที่มา : https://www.thaiprogrammer.org/2018/12อะไรคือ การเรียนรู้ของเครื่อง (Machine Learning)? (ฉบับมือใหม่)/?fbclid=IwAR1z3ALOzEAXzIbjCRql_aoMd25tpGwZ5QZbCfcFDA_XaAtu9FaZOb0ZVIo [01 เมษายน 2564]

ภาคผนวก

แบบเสนอหัวข้อโครงงาน รายวิชา 2301399 Project Proposal
ปีการศึกษา 2563

ชื่อโครงงาน (ภาษาไทย)	การประมาณค่าความหนาแน่นของ PM 2.5 ด้วยการเรียนรู้ของเครื่อง
ชื่อโครงงาน (ภาษาอังกฤษ)	A Density Estimation of PM 2.5 using Machine Learning
อาจารย์ที่ปรึกษา	รศ.ดร. ภัทรสินี ภัทรโกศล
ผู้ดำเนินการ	นายศีลวัต หมื่นสุวรรณ เลขประจำตัวนิสิต 6033539123
	สาขาวิชา คณิตศาสตร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
	คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

หลักการและเหตุผล

เนื่องด้วยปัจจุบัน PM 2.5 เป็นปัญหาอย่างนึงของประเทศไทยที่ถือว่าเป็นหนึ่งในปัญหาร้ายแรงในช่วงนี้ โดย PM ย่อมาจาก Particulate Matters คือ ค่ามาตรฐานของฝุ่นละอองที่ลอยในอากาศ ส่วน 2.5 คือ หน่วยวัดเป็นไมครอน ซึ่งเล็กกว่าเส้นผมมนุษย์หลายสิบเท่า จนไม่สามารถมองเห็นได้ด้วยตาเปล่า PM 2.5 เกิดจากการเผาไหม้ทั้งจากยานพาหนะ การเผาวัสดุการเกษตร ไฟป่า และกระบวนการอุตสาหกรรม PM 2.5 สามารถลอยละล่องอยู่ในอากาศเวียนไปวนมา เมื่อเราหายใจรับ PM 2.5 ผ่านเข้าจมูก ผ่านปาก หรือแม้แต่มาสัมผัสกับร่างกาย ทั้งเข้าตา ทั้งโดนใบหน้า มันสามารถเข้าไปถึงถุงลมในปอดได้ เป็นผลทำให้เกิดโรคในระบบทางเดินหายใจ และโรคปอดต่าง ๆ หากได้รับในปริมาณมากหรือเป็นระยะเวลานานจะสะสมในเนื้อเยื่อปอด ทำให้การทำงานของปอดเสื่อมประสิทธิภาพลง ทำให้หลอดลมอักเสบ มีอาการหอบหืด หรืออาจส่งผลอันตรายถึงชีวิต

PM 2.5 เริ่มเป็นที่รู้จักในคนไทยมากขึ้นในช่วง ธันวาคม 2560 เนื่องจากในช่วงนั้น PM 2.5 ได้มีปริมาณสูงขึ้นเป็นอย่างมากและส่งผลให้หน้ากากอนามัยมีราคาสูงขึ้นเป็นอย่างที่ไม่เคยเป็นมาก่อน รวมถึงทำให้เกิดค่าเสียโอกาสจากประเด็นด้านสุขภาพและการท่องเที่ยว รวมถึงค่าเสียโอกาสของภาคธุรกิจที่อาจได้รับผลกระทบจากการที่ประชาชนมีการปรับเปลี่ยนพฤติกรรม จึงก่อให้เกิดผลกระทบทางเศรษฐกิจ อีกทั้งประชาชนบางกลุ่มยังมีภาระในการใช้จ่ายด้านสุขภาพ และนักท่องเที่ยวต่างชาติเลือกที่จะปรับกิจกรรมการท่องเที่ยวของตน โดยอาจมีบางส่วนที่หลีกเลี่ยงการเดินทางเข้ามายังพื้นที่กรุงเทพมหานครและปริมณฑล

ด้วยเหตุนี้เอง รัฐบาลได้ออกมาตรการเพื่อใช้แก้ปัญหาที่เกิดขึ้น เช่น แผนปฏิบัติการขับเคลื่อนวาระแห่งชาติ การแก้ไขปัญหามลพิษด้านฝุ่นละอองขนาดเล็ก PM 2.5 การติดตาม

ตรวจสอบปริมาณฝุ่น PM2.5 และการดำเนินงานของกรมควบคุมมลพิษ มาตรการเพิ่มประสิทธิภาพในการบริหารจัดการเชิงพื้นที่ มาตรการและการป้องกันและลดการเกิดมลพิษที่ต้นทาง และมาตรการเพิ่มประสิทธิภาพการบริหารจัดการมลพิษ

ผู้วิจัยจึงเล็งเห็นความสำคัญของเรื่องนี้และต้องการที่จะรู้ว่า ปัจจัยใดมีผลต่อการเกิด PM 2.5 และต้องการที่จะพัฒนาให้เกิดเป็นโมเดลความสัมพันธ์ของปัจจัยที่พิจารณากับปริมาณ PM 2.5 ที่วัดได้ เพราะว่าผลกระทบของ PM 2.5 นั้นมีมากมาย ซึ่งหากเราสามารถคาดการณ์ปริมาณของ PM 2.5 ในอนาคตภายใต้ปัจจัยที่สามารถวัดได้ เราก็จะสามารถรับมือ แก้ไข และเตรียมพร้อมเพื่อลดความเสียหายที่จะเกิดขึ้นในด้านต่าง ๆ อันเกิดจาก PM 2.5 ได้

วัตถุประสงค์

1. กำหนดปัจจัยที่มีผลต่อการเกิด PM 2.5
2. พัฒนาโมเดลความสัมพันธ์ของปัจจัยที่พิจารณากับปริมาณ PM 2.5 ที่วัดได้

ขอบเขตของโครงงาน

1. ค่า PM 2.5 ที่จะนำมาพิจารณาจะอยู่ในช่วงปีพุทธศักราช 2563 เท่านั้น
2. แหล่งข้อมูลที่ใช้ คือ สำนักงานสถิติแห่งชาติ กรมอุตุนิยมวิทยา กรมควบคุมมลพิษ เว็บไซต์ Thai Meteorlogical Department เว็บไซต์ Air Quality Historical Data Platform และ Google Earth

วิธีการดำเนินงาน

1. ศึกษาหาความรู้เกี่ยวกับ PM 2.5 ปัจจัยต่าง ๆ ที่ส่งผลกระทบต่อ PM 2.5
2. ศึกษาวิธีใช้โปรแกรม RapidMiner Studio
3. เก็บรวบรวมข้อมูล PM 2.5 จากแหล่งข้อมูลต่าง ๆ
4. หาปัจจัยที่มีผลต่อการเกิด PM 2.5 โดยดูจากความสัมพันธ์ของค่านั้น ๆ กับ PM 2.5
5. หาโมเดลความสัมพันธ์เพื่อแสดงให้เห็นได้ว่า ปริมาณการเกิด PM 2.5 สามารถประมาณค่าได้จากปัจจัยต่าง ๆ ได้อย่างไร
6. สรุปผล และเขียนรายงาน

ขั้นตอน	เดือน									
	พ.ค.	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.
1.ศึกษาหาความรู้เกี่ยวกับ PM 2.5 ปัจจัยต่าง ๆ ที่ส่งผลกระทบต่อ PM 2.5										
2.ศึกษาวิธีใช้โปรแกรม SPSS										
3.เก็บรวบรวมข้อมูล PM 2.5 ข้อมูลทางอุตุนิยมวิทยา ข้อมูลทางธรณี										
4.หาปัจจัยที่มีผลต่อการเกิด PM 2.5 โดยดูจากความสัมพันธ์ของค่านั้น ๆ กับ PM 2.5										
5.พัฒนาโมเดลความสัมพันธ์										
6.สรุปผล และเขียนรายงาน										

ประโยชน์ที่คาดว่าจะได้รับ

สำหรับผู้ที่อ่านงานวิจัยนี้

1. ได้รู้จักกับ PM 2.5 มากขึ้น
2. รู้ว่าปัจจัยใดบ้างที่มีผลต่อ PM 2.5
3. ใช้โมเดลช่วยในการคาดการณ์ค่า PM 2.5 ในอนาคต เพื่อลดความเสียหายที่จะเกิดขึ้น
4. สามารถป้องกันตัวเองจาก PM 2.5

สำหรับผู้ทำงานวิจัยนี้

1. พัฒนาความสามารถในการคิด วิเคราะห์
2. เรียนรู้การวิเคราะห์ข้อมูล ปัจจัยใดที่ส่งผลกระทบกับประเด็นที่เราต้องการ
3. เรียนรู้การใช้งานโปรแกรม RapidMiner Studio
4. เรียนรู้วิธีการติดต่อกับหน่วยงานราชการ
5. มีประสบการณ์ในการทำงานจริง

อุปกรณ์และเครื่องมือที่ใช้

1. คอมพิวเตอร์
2. โปรแกรม RapidMiner Studio
3. โปรแกรม Microsoft office
4. วัสดุสำนักงาน

งบประมาณ

1.	MOUSE PAD	120	บาท
2.	KEYBOARD&MOUSE	1050	บาท
3.	SSD	1550	บาท
4.	RAM	1250	บาท
5.	หูฟัง	690	บาท
6.	กระดาษ A4	246	บาท
	รวม	4906	บาท

เอกสารอ้างอิง

1. BBC NEWS | ไทย. ฝุ่น PM 2.5 กลับมาวิกฤต รัฐบาลทำอะไรไปแล้วบ้างหลังประกาศเป็น “วาระแห่งชาติ” [ออนไลน์]. 2563. แหล่งที่มา : https://www.bbc.com/thai/thailand-51030236 [01 พฤศจิกายน 2563]
2. BBC NEWS | ไทย. ฝุ่น : เหตุใดสถานการณ์ฝุ่นละอองขนาดเล็ก จึงพุ่งสูงขึ้นมาอีกครั้ง [ออนไลน์]. 2561. แหล่งที่มา : https://www.bbc.com/thai/thailand-46643980 [01 พฤศจิกายน 2563]
3. Living sneak peek. PM 2.5 คืออะไร? [ออนไลน์]. 2563. แหล่งที่มา : https://www.livingsneakpeek.com/pm-2-5-คืออะไร/ [01 พฤศจิกายน 2563]
4. กรมควบคุมมลพิษ. ข้อมูลดัชนีคุณภาพอากาศ [ออนไลน์]. แหล่งที่มา : http://air4thai.pcd.go.th/webV2/aqi_info.php [01 พฤศจิกายน 2563]
5. กรมควบคุมมลพิษ. มาตรการป้องกันและแก้ไขปัญหาฝุ่นละออง PM 2.5 [ออนไลน์]. 2562. แหล่งที่มา : https://www.pdmo.go.th/pdmomedia/documents/2020/Jan/มาตรการป้องกันและแก้ไขปัญหาฝุ่นละออง%20PM%202.pdf [01 พฤศจิกายน 2563]

6. ประชาชาติธุรกิจ. ศูนย์วิจัยกสิกรไทยชี้ “ฝุ่น PM2.5 ที่รุนแรงขึ้น” ส่อกระทบเศรษฐกิจสูงถึง 6 พันล้านบาท [ออนไลน์]. 2563. แหล่งที่มา : https://www.prachachat.net/finance/news-413478 [01 พฤศจิกายน 2563]

ประวัติผู้เขียน

นาย ศีลวัต หมื่นสุวรรณ
รหัสนิสิต 6033539123
อีเมล์ ooviewoo@hotmail.com
สาขาวิชา คณิตศาสตร์
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย