



รายงานรายวิชา FRA 362 Robotics Studio V : Change the World Project
(Report for FRA 362 Robotics Studio V : Change the World Project)

Green Soil โมเดลทำนายคุณภาพดิน

สถาบันวิทยาการหุ่นยนต์ภาคสนาม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

กลุ่มที่ 8

ผู้ดำเนินงาน

1. นายภาวดล	กิรติพงษ์วุฒิ	รหัสนักศึกษา	65340500047
2. นางสาวศิริประภา	อุปภาค	รหัสนักศึกษา	65340500053
3. นายกษิดิษฐ์	รังสีวิจิตรประภา	รหัสนักศึกษา	65340500064
4. นางสาวชุตินันท์	พิพัฒนภิवงศ์	รหัสนักศึกษา	65340500068
5. นายสิปปวนนท์	สรณ์คุณแก้ว	รหัสนักศึกษา	65340500076

สารบัญ (1)

หัวข้อ	หน้า
สารบัญ	ก
สารบัญรูปภาพ.....	ก
สารบัญรูปภาพ (2).....	Error! Bookmark not defined.
สารบัญรูปภาพ (3).....	Error! Bookmark not defined.
สารบัญรูปภาพ (4).....	Error! Bookmark not defined.
สารบัญตาราง	ฉ
บทที่ 1 บทนำ	1
1.1 สถานการณ์ปัจจุบันของปัญหาดินเสื่อมสภาพ	1
1.2 ผลกระทบของการเสื่อมสภาพของดินต่อการเปลี่ยนแปลงสภาพภูมิอากาศ (Climate Change).....	3
1.3 การจัดการปัญหาดินเสื่อมสภาพ.....	6
บทที่ 2 แนวคิดนวัตกรรม	7
2.1 การสังเคราะห์และวิเคราะห์ข้อมูลปัญหา	7
2.2 System Map.....	8
2.3 แนวทางการแก้ไข	9
2.2.1 การสร้างโมเดล Machine Learning ทำนายคุณภาพดินในอนาคต	9
2.4 แนวคิดนวัตกรรม	10
บทที่ 3 วิธีการดำเนินงาน.....	11
3.1 รายละเอียดต้นแบบนวัตกรรม (Prototype)	11
3.1.1 ขอบเขตของต้นแบบนวัตกรรม	11
3.1.2 Requirement ของ Prototype.....	13

สารบัญ (2)

หัวข้อ	หน้า
3.2 รายละเอียดทางเทคนิค	15
3.2.1 Data Exploration	15
3.2.2 Data Preprocessing	38
3.2.3 Model Construction	50
บทที่ 4 ผลการทดสอบและอภิปราย	65
4.1 รายละเอียดการทดลอง	65
4.1.1 วัตถุประสงค์ของการทดลอง	65
4.1.2 ตัวแปรในการทดลอง	65
4.1.3 โมเดล Machine Learning ที่ใช้	65
4.1.4 การออกแบบการทดลอง	66
4.2 การทดสอบ	66
4.2.1 ขั้นตอนการทดลอง	66
4.2.2 การแสดงผลการทดลอง	66
4.3 การวิเคราะห์ผล	67
4.3.1 การวิเคราะห์ประสิทธิภาพของโมเดล	67
4.3.2 การวิเคราะห์ความสัมพันธ์ของตัวแปร	67
4.3.3 ประเด็นที่ค้นพบจากการทดลอง	67
4.3.4 เปรียบเทียบระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012	67
บทที่ 5 บทสรุปการดำเนินงานและข้อเสนอแนะ	74
5.1 สรุปการดำเนินงาน	74
5.2 ความสามารถและข้อจำกัดของ Prototype	74

สารบัญ (3)

หัวข้อ	หน้า
5.3 ปัญหาที่พบในการดำเนินงาน	76
5.4 แนวทางการพัฒนาในอนาคต	76
อ้างอิง	78

สารบัญรูปภาพ (1)

หัวข้อ	หน้า
รูป 1 สถานการณ์ปัญหาเรื่องดินเสื่อมสภาพในปี 2015.....	1
รูป 2 สถานการณ์ปัญหาเรื่องดินเสื่อมสภาพในปี 2019.....	2
รูป 3 ปริมาณการปล่อยก๊าซเรือนกระจกจำแนกตามภาคส่วนตั้งแต่ปีพ.ศ. 2543 - พ.ศ. 2562	3
รูป 4 การเปรียบปริมาณการปล่อยก๊าซเรือนกระจกระหว่างพ.ศ. 2543 และ พ.ศ. 2562	4
รูป 5 กิจกรรมด้านการเกษตรที่ส่งผลต่อการปล่อยก๊าซเรือนกระจก.....	4
รูป 6 System Map.....	8
รูป 7 แผนภาพการทดลองโดยรวม	12
รูป 8 แสดงประเภทข้อมูลของ Target.....	20
รูป 9 แสดงประเภทข้อมูลของชุดข้อมูลปี 2009.....	21
รูป 10 แสดงประเภทข้อมูลของชุดข้อมูลปี 2010	22
รูป 11 แสดงประเภทข้อมูลของชุดข้อมูลปี 2011.....	23
รูป 12 การแสดงการกระจายตัวข้อมูลของ target ในปี 2012 (1).....	24
รูป 13 การแสดงการกระจายตัวข้อมูลของ target ในปี 2012 (2).....	25
รูป 14 แสดงการกระจายตัวของชุดข้อมูลปี 2009 (1).....	26
รูป 15 แสดงการกระจายตัวของชุดข้อมูลปี 2009 (2).....	27
รูป 16 แสดงการกระจายตัวของชุดข้อมูลปี 2009 (3).....	28
รูป 17 แสดงการกระจายตัวของชุดข้อมูลปี 2009 (4).....	29
รูป 18 แสดงการกระจายชุดข้อมูลปี 2010 (1)	30
รูป 19 แสดงการกระจายชุดข้อมูลปี 2010 (2)	31
รูป 20 แสดงการกระจายชุดข้อมูลปี 2010 (3)	32
รูป 21 แสดงการกระจายชุดข้อมูลปี 2010 (4)	33
รูป 22 แสดงการกระจายชุดข้อมูลปี 2011(1).....	34
รูป 23 แสดงการกระจายชุดข้อมูลปี 2011(2).....	35
รูป 24 แสดงการกระจายชุดข้อมูลปี 2011(3).....	36
รูป 25 แสดงการกระจายชุดข้อมูลปี 2011(4).....	37

สารบัญรูปภาพ (2)

หัวข้อ	หน้า
รูป 26 missing value ทุกปี.....	38
รูป 27 missing value ของแต่ละชุดข้อมูลหลังจากไส้ค่าเฉลี่ยแต่ละชุดข้อมูล	38
รูป 28 การกระจายตัวของข้อมูล ปี 2012 หลังการจัดการข้อมูลผิดปกติ (1)	39
รูป 29 การกระจายตัวของข้อมูล ปี 2012 หลังการจัดการข้อมูลผิดปกติ (2)	40
รูป 30 การกระจายตัวของข้อมูล ปี 2012 หลังการจัดการข้อมูลผิดปกติ (3)	40
รูป 31 การกระจายตัวของข้อมูล ปี 2012 หลังการจัดการข้อมูลผิดปกติ (4)	41
รูป 32 การกระจายตัวของข้อมูล ปี 2009 หลังการจัดการข้อมูลผิดปกติ (1)	42
รูป 33 การกระจายตัวของข้อมูล ปี 2009 หลังการจัดการข้อมูลผิดปกติ (2)	42
รูป 34 การกระจายตัวของข้อมูล ปี 2009 หลังการจัดการข้อมูลผิดปกติ (3)	43
รูป 35 การกระจายตัวของข้อมูล ปี 2009 หลังการจัดการข้อมูลผิดปกติ (4)	43
รูป 36 การกระจายตัวของข้อมูล ปี 2010 หลังการจัดการข้อมูลผิดปกติ (1)	44
รูป 37 การกระจายตัวของข้อมูล ปี 2010 หลังการจัดการข้อมูลผิดปกติ (2)	44
รูป 38 การกระจายตัวของข้อมูล ปี 2010 หลังการจัดการข้อมูลผิดปกติ (3)	45
รูป 39 การกระจายตัวของข้อมูล ปี 2011 หลังการจัดการข้อมูลผิดปกติ (4)	45
รูป 40 โปรแกรมการใช้ StandardScaler(Standardization).....	50
รูป 41 ลักษณะการแบ่งข้อมูลในการเทรนโมเดล	50
รูป 42 แสดงหน้าต่างระบุชื่อ Target, หน่วย, ค่าเฉลี่ยรายปี และผลการประเมิน	52
รูป 43 แสดงค่าของข้อมูล C % Inorg ที่มีการระบุช่วงที่เป็นผลดีต่อдин	52
รูป 44 แสดงค่าของข้อมูล C % Org ที่มีการระบุช่วงที่เป็นผลดีต่อдин	53
รูป 45 แสดงค่าของข้อมูล EC ที่มีการระบุช่วงที่เป็นผลดีต่อдин	53
รูป 46 แสดงค่าของข้อมูล ECEC ที่มีการระบุช่วงที่เป็นผลดีต่อдин	54
รูป 47 แสดงค่าของข้อมูล Ksat ที่มีการระบุช่วงที่เป็นผลดีต่อдин	54
รูป 48 แสดงค่าของข้อมูล M3 Al ที่มีการระบุช่วงที่เป็นผลดีต่อдин	55
รูป 49 แสดงค่าของข้อมูล M3 B ที่มีการระบุช่วงที่เป็นผลดีต่อдин	55
รูป 50 แสดงค่าของข้อมูล M3 Ca ที่มีการระบุช่วงที่เป็นผลดีต่อдин	56

สารบัญรูปภาพ (3)

หัวข้อ	หน้า
รูป 51 แสดงค่าของข้อมูล M3 Cu ที่มีการระบุช่วงที่เป็นผลดีต่อдин	56
รูป 52 แสดงค่าของข้อมูล M3 Fe ที่มีการระบุช่วงที่เป็นผลดีต่อдин	57
รูป 53 แสดงค่าของข้อมูล M3 K ที่มีการระบุช่วงที่เป็นผลดีต่อдин	57
รูป 54 แสดงค่าของข้อมูล M3 Mg ที่มีการระบุช่วงที่เป็นผลดีต่อдин	58
รูป 55 แสดงค่าของข้อมูล M3 Mn ที่มีการระบุช่วงที่เป็นผลดีต่อдин	58
รูป 56 แสดงค่าของข้อมูล M3 Na ที่มีการระบุช่วงที่เป็นผลดีต่อдин	59
รูป 57 แสดงค่าของข้อมูล M3 P ที่มีการระบุช่วงที่เป็นผลดีต่อдин	59
รูป 58 แสดงค่าของข้อมูล M3 S ที่มีการระบุช่วงที่เป็นผลดีต่อдин	60
รูป 59 แสดงค่าของข้อมูล M3 Zn ที่มีการระบุช่วงที่เป็นผลดีต่อдин	60
รูป 60 แสดงค่าของข้อมูล PH ที่มีการระบุช่วงที่เป็นผลดีต่อдин	61
รูป 61 แสดงค่าของข้อมูล Psa aclay ที่มีการระบุช่วงที่เป็นผลดีต่อдин	61
รูป 62 แสดงค่าของข้อมูล Psa asand ที่มีการระบุช่วงที่เป็นผลดีต่อдин	62
รูป 63 แสดงค่าของข้อมูล Psa asilt ที่มีการระบุช่วงที่เป็นผลดีต่อдин	62
รูป 64 แสดงค่าของข้อมูล Total carbon ที่มีการระบุช่วงที่เป็นผลดีต่อдин	63
รูป 65 แสดงค่าของข้อมูล Total nitrogen ที่มีการระบุช่วงที่เป็นผลดีต่อдин	63
รูป 66 แสดงค่าของข้อมูล Wcvfrsat ที่มีการระบุช่วงที่เป็นผลดีต่อдин	64
รูป 67 Bar Plot เปรียบเทียบระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (random forest)	68
รูป 68 การเปรียบเทียบความแม่นยำระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (random forest)	68
รูป 69 Bar Plot เปรียบเทียบระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (SVR)	69
รูป 70 การเปรียบเทียบความแม่นยำระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (SVR)	69
รูป 71 Bar Plot เปรียบเทียบระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (Decision Tree)	70
รูป 72 การเปรียบเทียบความแม่นยำระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (Decision Tree)	70
รูป 73 กราฟ Bar Plot เปรียบเทียบระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (Gradient Boosting)	71
รูป 74 การเปรียบเทียบความแม่นยำระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (Gradient Boosting)	71
รูป 75 Bar Plot เปรียบเทียบระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (Extreme Gradient Boosting)	72

สารบัญรูปภาพ (4)

หัวข้อ	หน้า
รูป 76 เปรียบเทียบความแม่นยำระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (Extreme Gradient Boosting)	72
รูป 77 Bar Plot เปรียบเทียบระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (Neural Network).....	73
รูป 78 การเปรียบเทียบความแม่นยำระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (Neural Network)	73
รูป 79 แนวโน้มของ M3 AI ของทั้ง 4 ปี.....	76

สารบัญตาราง

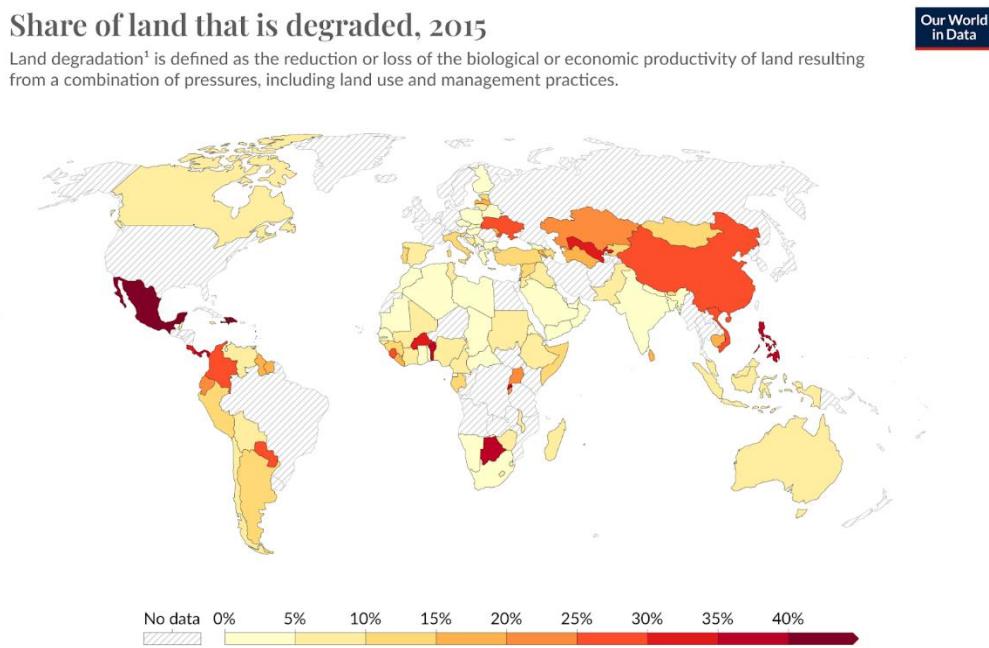
หัวข้อ	หน้า
ตาราง 1 Technological Feasibility.....	9
ตาราง 2 ความหมายของ feature.....	15
ตาราง 3 ข้อมูลที่ใช้ในการคาดการณ์.....	18
ตาราง 4 ชุดข้อมูลที่มีความสัมพันธ์กับ target.....	46
ตาราง 5 การเปรียบเทียบโมเดล	51

บทที่ 1 บทนำ

การเสื่อมคุณภาพของดิน

ดินเป็นทรัพยากรธรรมชาติที่มีบทบาทสำคัญในการสนับสนุนชีวิตและระบบ生นิเวศทั่วโลก โดยทำหน้าที่สำคัญ ทั้งการเจริญเติบโตของพืช การกักเก็บคาร์บอน และการหมุนเวียนธาตุอาหาร ฯลฯ อย่างไรก็ตาม องค์การอาหารและการเกษตรแห่งสหประชาชาติ (FAO) ระบุว่า 40% ของพื้นที่ดินทั่วโลกกำลังเผชิญกับภาวะเสื่อมโทรมอย่างรุนแรงสถานการณ์ดังกล่าวเด่นชัดในพื้นที่เกษตรกรรมที่มีการจัดการดินไม่เหมาะสม เช่น การใช้สารเคมีเกินความจำเป็นและการพัฒนาอย่างดินในพื้นที่ลาดชัน จากปัญหานี้ก่อให้เกิดการลดศักยภาพของดินในการกักเก็บคาร์บอนและเร่งการปลดปล่อยคาร์บอนไดออกไซด์ ซึ่งเป็นปัจจัยที่ทำให้การเปลี่ยนแปลงสภาพภูมิอากาศทวีความรุนแรง (UNCCD, 2023)

1.1 สถานการณ์ปัจจุบันของปัญหาดินเสื่อมสภาพ

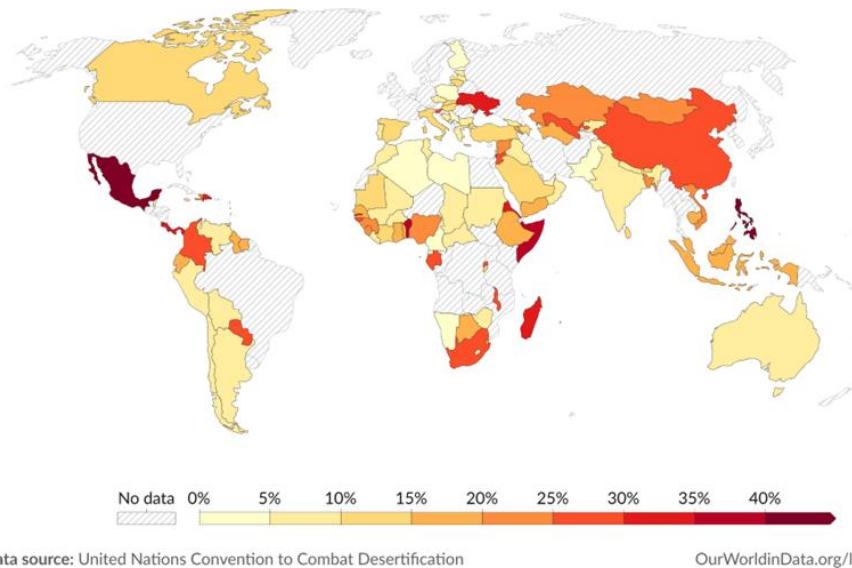


รูป 1 สถานการณ์ปัญหารှိองดินเสื่อมสภาพในปี 2015

Share of land that is degraded, 2019

Our World
in Data

Land degradation¹ is defined as the reduction or loss of the biological or economic productivity of land resulting from a combination of pressures, including land use and management practices.



รูป 2 สถานการณ์ปัญหารှောက်ดินเสื่อมสภาพในปี 2019

แผนที่โลกนี้แสดงถึงการเปลี่ยนแปลงของพื้นที่ดินเสื่อมโทรมระหว่างปี 2015 และ 2019 ซึ่งสืบให้เห็นถึงการเพิ่มขึ้นของพื้นที่ที่เสื่อมคุณภาพอย่างชัดเจน โดยในปี 2015 (พ.ศ. 2558) มีพื้นที่ที่เกิดปัญหาดินเสื่อมโทรมรุนแรง (สีแดงเข้ม) ในภูมิภาคแอฟริกาตะวันออก เช่น โซมาเลีย และบางส่วนของเอเชียกลาง รวมถึงภูมิภาคอเมริกากลาง เช่น เม็กซิโก โดยดินในพื้นที่เหล่านี้ประสบปัญหาสูญเสียความอุดมสมบูรณ์ ส่งผลกระทบต่อการเกษตรและระบบนิเวศโดยรวม

เมื่อเปรียบเทียบกับปี 2019 (ค.ศ. 2562) พบว่าพื้นที่ที่มีปัญหาดินเสื่อมโทรมได้ขยายตัวมากขึ้นอย่างชัดเจน โดยเฉพาะในภูมิภาคเอเชียตะวันออกเฉียงใต้ รวมถึงยุโรปตะวันออก ซึ่งการเปลี่ยนแปลงนี้สะท้อนถึงผลกระทบของการใช้ทรัพยากรอย่างหนักในพื้นที่เกษตรกรรม การจัดการดินที่ไม่เหมาะสมต่อการใช้งาน เช่น การใช้สารเคมีในปริมาณมาก และการตัดไม้ทำลายป่าในบางภูมิภาค

ในเชิงสถิติ พบว่าพื้นที่ดินเสื่อมโทรมทั่วโลกเพิ่มขึ้นประมาณ 15% ในช่วงเวลาหนึ่ง โดยเฉพาะพื้นที่เกษตรกรรมในเอเชียตะวันออกเฉียงใต้ที่มีอัตราการเพิ่มขึ้นสูงถึง 18% ในบางประเทศ ขณะที่ในยุโรปตะวันออกพื้นที่สีแดงเข้มเพิ่มขึ้นประมาณ 12% ซึ่งสะท้อนถึงความจำเป็นเร่งด่วนในการพัฒนากลยุทธ์การจัดการดินอย่างยั่งยืนเพื่อรับปัญหาที่เพิ่มขึ้น

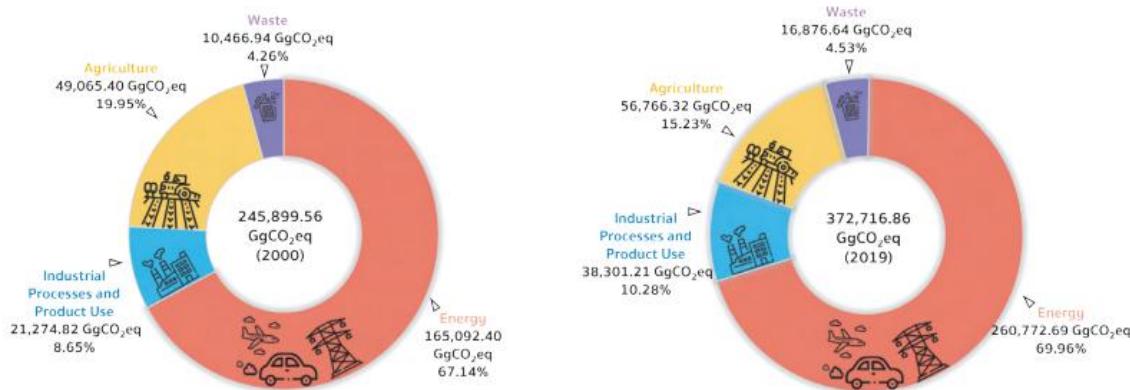
1.2 ผลกระทบของการเสื่อมสภาพของดินต่อการเปลี่ยนแปลงสภาพภูมิอากาศ (Climate Change)

Unit: GgCO₂eq

Year	Source category					Net emissions (Include LULUCF)	Total emissions (Exclude LULUCF)
	Energy	Industrial Processes and Product Use	Agriculture	LULUCF	Waste		
2000	165,092.40	21,274.82	49,065.40	-45,443.60	10,466.94	200,455.96	245,899.56
2001	172,906.76	22,873.13	50,019.89	-54,436.91	11,281.28	202,644.15	257,081.06
2002	182,317.06	24,944.60	48,691.69	-44,126.05	12,813.68	224,640.98	268,767.03
2003	191,305.31	24,217.55	52,265.60	-45,887.41	14,281.85	236,182.90	282,070.31
2004	208,180.48	26,041.51	51,958.87	-54,146.05	14,844.36	246,879.17	301,025.22
2005	212,318.12	27,883.81	51,413.70	-58,340.43	15,167.72	248,442.92	306,783.35
2006	213,020.12	28,771.00	53,870.52	-61,390.87	16,188.61	250,459.38	311,850.25
2007	219,378.63	30,424.56	57,161.74	-68,396.37	16,569.69	255,138.25	323,534.62
2008	221,716.62	28,897.09	59,059.43	-69,184.97	17,652.74	258,140.91	327,325.88
2009	221,351.88	28,344.22	61,137.46	-71,197.04	17,455.59	257,092.11	328,289.15
2010	233,105.75	29,961.53	60,694.70	-65,827.73	15,778.55	273,712.80	339,540.53
2011	229,859.44	31,290.10	61,220.57	-74,636.18	14,900.97	262,634.90	337,271.08
2012	242,867.49	33,484.60	64,250.78	-80,695.94	11,866.89	271,773.82	352,469.76
2013	241,717.77	33,299.85	59,080.05	-83,110.84	13,978.13	264,964.96	348,075.80
2014	239,724.21	34,819.39	57,995.56	-100,508.25	14,237.77	246,268.68	346,776.93
2015	241,918.80	35,675.42	52,324.55	-87,337.57	15,998.67	258,579.87	345,917.44
2016	254,540.07	37,189.17	52,303.38	-87,054.60	16,771.89	273,749.91	360,804.51
2017	255,244.90	38,054.69	57,052.83	-85,380.82	17,666.21	282,637.93	368,018.75
2018	257,340.89	40,118.18	58,486.02	-85,968.30	16,703.68	286,680.47	372,648.77
2019	260,772.69	38,301.21	56,766.32	-91,988.52	16,876.64	280,728.34	372,716.86

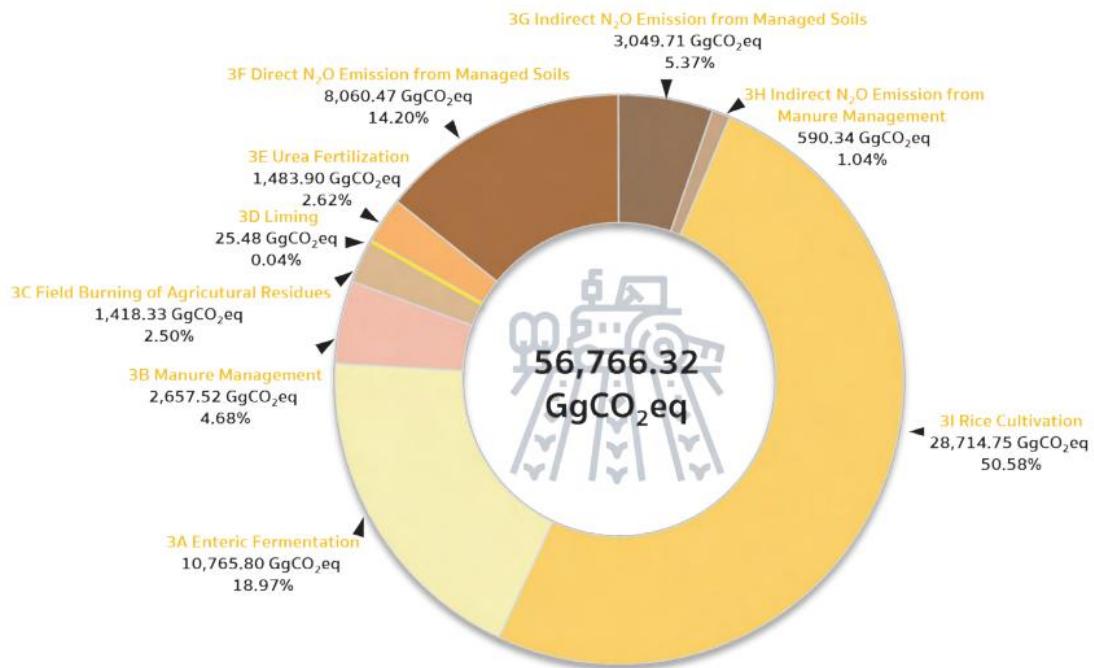
รูป 3 ปริมาณการปล่อยก๊าซเรือนกระจกจำกัดแนบทามภาคส่วนตั้งแต่ปีพ.ศ. 2543 - พ.ศ. 2562

ที่มา : <https://eservice.dcce.go.th/storage/Media/C202401185676.pdf>



รูป 4 การเปรียบปริมาณการปล่อยก๊าซเรือนกระจกระหว่าง พ.ศ. 2543 และ พ.ศ. 2562

ที่มา : <https://eservice.dcce.go.th/storage/Media/C202401185676.pdf>



รูป 5 กิจกรรมด้านการเกษตรที่ส่งผลต่อการปล่อยก๊าซเรือนกระจก

ที่มา : <https://eservice.dcce.go.th/storage/Media/C202401185676.pdf>

การเสื่อมสภาพของดินเป็นปัจจัยสำคัญที่ส่งผลต่อการเปลี่ยนแปลงสภาพภูมิอากาศ เนื่องจากดินที่เสื่อมสภาพมีความสามารถในการกักเก็บคาร์บอนลดลง ซึ่งนำไปสู่การปล่อยก๊าซเรือนกระจกสู่ชั้นบรรยากาศเพิ่มขึ้น โดยเฉพาะก๊าซcarbon dioxide (CO_2) และไนตรัสออกไซด์ (N_2O)

จากสถิติในปี 2019 การปล่อยก๊าซเรือนกระจกจากการเกษตรกรรมคิดเป็น 15.23% หรือประมาณ 56,766.32 $\text{Gg CO}_2\text{eq}$ ของการปล่อยทั้งหมด โดยเกิดจากการทำนาข้าว การใช้ปุ๋ยในโตรjenที่มากเกินไปและการจัดการดินที่ไม่เหมาะสม จนกลายเป็นการเสื่อมสภาพของดินที่ส่งผลต่อความอุดมสมบูรณ์ของดิน และเร่งให้เกิดการปล่อยก๊าซเรือนกระจก และนำไปสู่การเปลี่ยนแปลงสภาพภูมิอากาศ โดยผลกระทบหลักที่เกี่ยวข้องกับดินเสื่อมสภาพ ได้แก่

1.2.1 การปล่อยก๊าซเรือนกระจก

คาร์บอนไดออกไซด์ (CO_2) และไนตรัสออกไซด์ (N_2O) ที่เกิดจากการสลายตัวของอินทรีย์วัตถุ ในดินที่ไม่ได้ถูกกักเก็บไว้ในสภาพดินที่เหมาะสม โดยเมื่อดินมีการสูญเสียโครงสร้างและความอุดมสมบูรณ์ภายในดิน จะทำให้เกิดการปล่อย CO_2 สูงขึ้นประมาณ 20% เนื่องจากดินไม่สามารถกักเก็บคาร์บอนได้อย่างมีประสิทธิภาพ ส่งผลให้ก๊าซเรือนกระจกสะสมในชั้นบรรยากาศเพิ่มมากขึ้น

1.2.2. ความทันทานของดินต่อสภาพอากาศสุดขั้วลดลง

ดินที่มีความอุดมสมบูรณ์ต่ำไม่สามารถทนทานต่อสภาพอากาศสุดขั้ว เช่น ภาวะแห้งแล้งหรือฝนตกหนักได้ โดยภาวะแห้งแล้งทำให้ดินสูญเสียความชุ่มชื้นและโครงสร้างดินแข็งตัว เกิดการแตกร้าวนยกต่อการดูดซับน้ำ ในขณะที่ฝนตกหนักสามารถละล้างธาตุอาหารและโครงสร้างดินออกไป ส่งผลให้เกิดการพังทลายของดินในพื้นที่ลาดชัน รวมถึงน้ำท่วมขังในพื้นที่ราบต่ำ นอกจากนี้ โครงสร้างดินที่เสื่อมสภาพยังลดความสามารถในการกักเก็บคาร์บอนและการดูดซับก๊าซเรือนกระจก เช่น คาร์บอนไดออกไซด์และไนตรัสออกไซด์ ซึ่งเป็นปัจจัยที่เร่งกระบวนการเปลี่ยนแปลงสภาพภูมิอากาศให้รุนแรงยิ่งขึ้น

โดยการเสื่อมโทรมของดินทั่วโลกมีอัตราเพิ่มขึ้นประมาณ 15% ในช่วงสองทศวรรษที่ผ่านมา และคาดการณ์ว่าภายในปี 2050 ปัญหานี้จะกระทบพื้นที่เกษตรกรรมที่สำคัญในแอฟริกาและเอเชียตะวันออกเฉียงใต้โดยตรง (UNCCD, 2023)

1.3 การจัดการปัญหาดินเสื่อมสภาพ

การเสื่อมโทรมของดินมักเกิดขึ้นอย่างช้า ๆ และไม่มีสัญญาณเตือนที่ชัดเจน ทำให้ผู้คนส่วนมากจะตระหนักรถึงปัญหาที่ต่อเมื่อติดสูญเสียความอุดมสมบูรณ์ไปจนถึงดินสูญเสียคุณสมบัติของดินจนยากจะแก้ไขได้

1.3.1 วิธีการแก้ไขปัญหาดินเสื่อมคุณภาพในปัจจุบัน

1. การใช้ปุ๋ยเคมีและปุ๋ยอินทรีย์

การเติมธาตุอาหารลงในดินช่วยเพิ่มความอุดมสมบูรณ์ในระยะสั้น แต่ไม่สามารถแก้ไขปัญหาโครงสร้างดินที่เสียหายได้ นอกจากนี้ การใช้ปุ๋ยเคมีในปริมาณมากยังส่งผลกระทบต่อสิ่งแวดล้อม เช่น การปนเปื้อนในน้ำได้ดี

2. การปลูกพืชคลุมดิน

วิธีนี้ช่วยลดการพังทลายของดินและเพิ่มอินทรีย์วัตถุในดิน แต่ในพื้นที่ที่ดินเสื่อมโทรมรุนแรง การฟื้นฟูด้วยวิธีนี้ต้องใช้เวลาและทรัพยากรจำนวนมาก

3. การปรับปรุงระบบชลประทาน

การจัดการน้ำอย่างเหมาะสมช่วยลดการชะล้างหน้าดินและธาตุอาหาร แต่ต้องการการลงทุนที่สูงและการบำรุงรักษาอย่างต่อเนื่อง ซึ่งไม่เหมาะสมสำหรับเกษตรกรรายย่อยในหลายพื้นที่

1.3.2 ข้อจำกัดในการแก้ไขปัญหาระดับดินเสื่อมสภาพ

วิธีการแก้ไขปัญหาดินเสื่อมคุณภาพในปัจจุบันที่มีอยู่สามารถช่วยบรรเทาผลกระทบในระยะสั้นได้ในบางกรณี แต่โดยส่วนใหญ่จะมุ่งเน้นไปที่การแก้ไขปัญหาปลายเหตุ แทนที่จะให้ความสำคัญกับการป้องกันปัญหาในระยะยาว ส่งผลให้การจัดการปัญหาไม่ได้รับการดำเนินการอย่างยั่งยืน นอกจากนี้ การฟื้นฟูดินที่เสื่อมโทรมยังต้องอาศัยทรัพยากรในหลายด้าน ไม่ว่าจะเป็นทั้งทางด้านของเวลา ด้านเงินทุน หรือความเชี่ยวชาญ ซึ่งในบางพื้นที่ที่มีข้อจำกัดด้านทรัพยากร ทำให้ไม่สามารถดำเนินการฟื้นฟูได้อย่างประสิทธิภาพเท่าที่ควร

บทที่ 2 แนวคิดนวัตกรรม

2.1 การสังเคราะห์และวิเคราะห์ข้อมูลปัญหา

ปัญหาดินเสื่อมสภาพส่งผลกระทบอย่างกว้างขวางต่อทั้งระบบการเกษตร ความมั่นคงทางอาหาร และระบบนิเวศ โดยข้อมูลปัญหาที่ได้ชี้ให้เห็นถึงปัจจัยที่เกี่ยวข้องดังนี้

2.1.1 สาเหตุของปัญหา

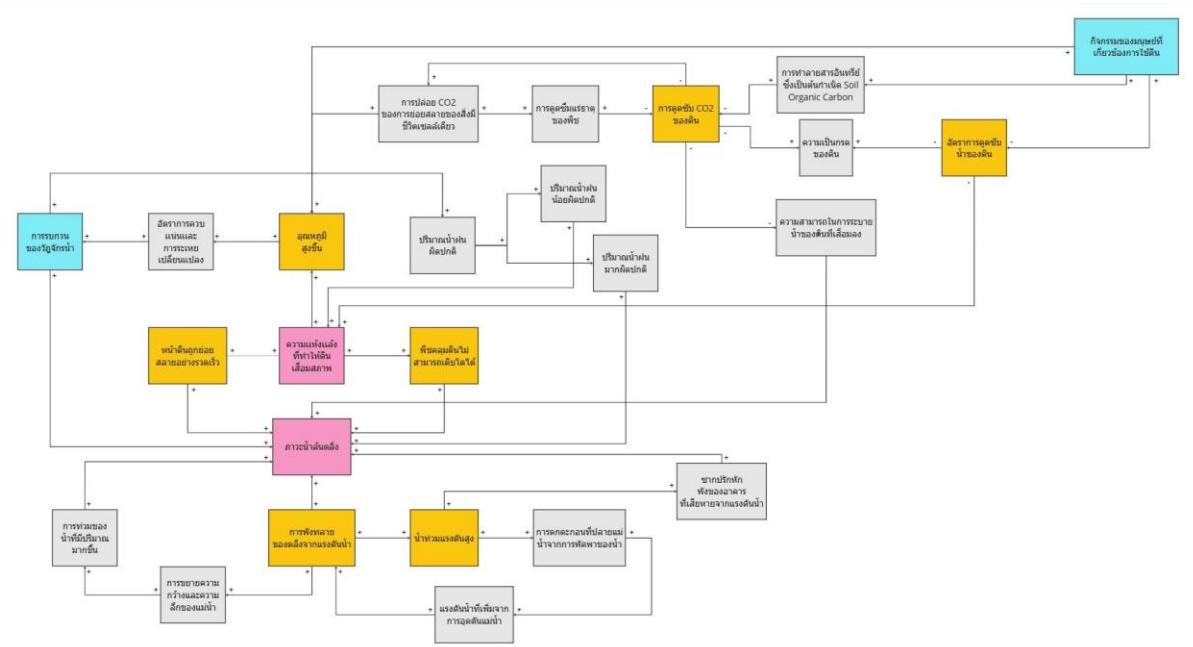
- การจัดการปุ๋ยและการเกษตรที่ไม่เหมาะสม เช่น การใช้สารเคมีเกินความจำเป็น
- การสูญเสียความอุดมสมบูรณ์ของดิน เนื่องจากขาดแคลนสารอินทรีย์ เช่น คาร์บอนในดิน
- การเปลี่ยนแปลงการใช้พื้นที่ดิน เช่น การตัดไม้ทำลายป่าและการขยายพื้นที่เกษตร
- การชะล้างและพังทลายของดิน โดยเฉพาะในพื้นที่ลาดชัน
- การสะสมของเกลือในดิน (Salinization) ซึ่งมักเกิดจากการฉลุประทานที่ไม่เหมาะสม

2.1.2 ผลกระทบ

- ลดผลผลิตทางการเกษตร ส่งผลกระทบต่อความมั่นคงทางอาหารและรายได้ของเกษตรกร
- ความเสียหายต่อระบบนิเวศ โดยเฉพาะการสูญเสียความสามารถในการกักเก็บcarbon
- ความยั่งยืนของระบบการเกษตรลดลง ทำให้เกิดปัญหาความมั่นคงทางอาหารในระยะยาว

2.2 System Map

System Map ด้านล่างนี้แสดงถึงการเขื่อมโยงปัจจัยต่าง ๆ ของปัญหาดินสู่อุณหภูมิโลก พร้อมทั้งจุดอ่อน (Weak Points) ที่สามารถแก้ปัญหาได้



รูป 6 System Map

โดย สีฟ้า คือ ถ้าไม่เกิดปรากฏการณ์นี้จะไม่เกิดเหตุการณ์ทั้งหมด

สีเหลือง คือ action ที่เกิดจากธรรมชาติหรือมนุษย์

สีขาว คือ ปรากฏการณ์ที่ส่งผลให้เกิดปัญหา Climate Change (สีชมพู)

สีชมพู คือ ปัญหา Climate Change ที่เราสนใจ

2.2.1 Key Weak Points:

1. การจัดการดินแบบตั้งเดิม

- เกษตรกรบางกลุ่มยังขาดความรู้ในการจัดการดินที่เหมาะสม

2. ข้อมูลที่ไม่เพียงพอ

- การขาดข้อมูลเชิงพื้นที่ เช่น pH, ความชื้น หรือปริมาณสารอาหารในดิน

- เกษตรกรหรือคนในพื้นที่ไม่สามารถรับรู้สถานการณ์การเสื่อมสภาพของดินล่วงหน้าได้ หรือ ถ้าสามารถรับรู้ได้ ก็ต่อเมื่อดินมีความเสื่อมสภาพมากแล้ว
3. ต้นทุนการปรับปรุงสูง
- เทคโนโลยีบางประเภท เช่น การใช้ปุ๋ยอินทรีย์ อาจมีราคาสูง

2.3 แนวทางการแก้ไข

จากการวิเคราะห์ใน System Map การทราบสถานะคุณภาพของดินอย่างต่อเนื่องและแม่นยำก่อนที่จะเกิดปัญหาเป็นสิ่งจำเป็น ระบบการตรวจสอบและวิเคราะห์คุณภาพดินที่ทันสมัย จึงต้องสามารถให้ข้อมูลเชิงลึก รวมถึงแนะนำแนวทางการจัดการที่เหมาะสมให้แก่เกษตรกรได้อย่างทันท่วงที กลุ่มของเราได้ออกแบบแนวทางการทำน้ำตกรรมโดยใช้เทคโนโลยี Machine Learning (ML) และ Internet of Things (IoT) ทำให้ทางกลุ่มคิดแนวทางการแก้ไขที่สามารถพัฒนาได้ดังนี้

2.2.1 การสร้างโมเดล Machine Learning นำทางคุณภาพดินในอนาคต

- ใช้ข้อมูลเชิงพื้นที่และข้อมูลจากเซ็นเซอร์ เช่น pH, ความชื้น, และระดับสารอาหารในดิน
- สร้างโมเดลการทำนายคุณภาพดินในอนาคต พร้อมแนะนำแนวทางการจัดการอย่างเหมาะสม

2.2.2 ระบบตรวจสอบและวิเคราะห์คุณภาพดินด้วย IoT

- การใช้เซ็นเซอร์ IoT ติดตามข้อมูลแบบเรียลไทม์ เช่น อุณหภูมิและความชื้น
- การใช้ AI วิเคราะห์ข้อมูลเพื่อให้คำแนะนำที่เหมาะสมต่อเกษตรกร

ตาราง 1 Technological Feasibility

Technological Feasibility		Economic Feasibility	Legal Feasibility	Operational Feasibility		Schedule Feasibility	Total Feasibility Score	Troubleshooting Guidelines
Is the solution technically possible?	Is our team capable of utilizing the technology?	Is the project financially attractive?	Is the project legal?	Does our team have the existing environment to support the project?	How well is the solution going to solve the problem?	Can the project be done in time?		
4	4	5	4	3	4	2	26	ML
4	4	4	4	3	4	3	26	IoT

2.4 แนวคิดนวัตกรรม

จัดทำได้เลือกพัฒนา โมเดล Machine Learning นำมายกคุณภาพดินในอนาคต สำหรับตรวจสอบและคาดการณ์คุณภาพดินในอนาคต เพื่อช่วยในการตัดสินใจและวางแผน การจัดการดินอย่างเหมาะสม เนื่องจากปัญหาดินเสื่อมสภาพเป็นปัญหาที่ส่งผลกระทบต่อผลผลิตทางการเกษตร ความมั่นคงทางอาหาร และสิ่งแวดล้อม โดยมีสาเหตุมาจากการใช้งานดินที่ไม่เหมาะสม เช่น การใช้ปุ๋ยเคมีเกินขนาด และการฉะล้างหน้าดิน

นวัตกรรมดังกล่าวช่วยวิเคราะห์ดินและคาดการณ์คุณภาพดินในอนาคตเพื่อช่วยส่งเสริมการจัดการดินอย่างยั่งยืน โดยมีบทบาทสำคัญในการลดผลกระทบจาก Climate Change เช่น การเพิ่มศักยภาพดินในการกักเก็บคาร์บอน และลดการปลดปล่อยก๊าซเรือนกระจก ทั้งนี้ ระบบนี้ยังสนับสนุนการทำเกษตรกรรมที่เป็นมิตรต่อสิ่งแวดล้อม และเป็นแนวทางสำคัญในการพัฒนาอย่างยั่งยืนในอนาคต

ระบบนี้มุ่งเน้นการวิเคราะห์คุณภาพดินโดยใช้เทคโนโลยี Machine Learning เพื่อช่วยในการประมาณผลข้อมูลที่เกี่ยวข้องกับคุณภาพดิน เช่น ค่า pH ความชื้นในดิน อุณหภูมิ ปริมาณอินทรีย์วัตถุในดิน และปริมาณธาตุอาหารหลักอย่างไนโตรเจน ฟอสฟอรัส และโพแทสเซียม ข้อมูลดังกล่าวจะถูกนำมาวิเคราะห์เชิงลึกเพื่อทำนายแนวโน้มในอนาคต

โมเดล Machine Learning ถูกออกแบบให้สามารถตรวจจับความผิดปกติของดิน เช่น การลดลงของอินทรีย์วัตถุ หรือการเปลี่ยนแปลงค่า pH ที่อาจส่งผลกระทบต่อการเจริญเติบโตของพืช โมเดลที่สนใจมีทั้งหมด 6 ประเภท ได้แก่ Random Forest, Support Vector Regression (SVM), Decision Tree, Gradient Boosting, Extreme Gradient Boosting และ Neural Networks

เมื่อการประมาณผลเสร็จสมบูรณ์ ระบบจะส่งการแจ้งเตือนถึงผู้ใช้งานเกี่ยวกับการเปลี่ยนแปลงของคุณภาพดินที่อาจเกิดขึ้นในอนาคต เพื่อช่วยในการตัดสินใจและวางแผนการจัดการพื้นที่เพาะปลูกได้อย่างเหมาะสม

บทที่ 3 วิธีการดำเนินงาน

3.1 รายละเอียดต้นแบบนวัตกรรม (Prototype)

จากแนวคิดนวัตกรรมระบบวิเคราะห์คุณภาพดินโดยใช้เทคโนโลยี Machine Learning ที่อธิบายไว้ในบทที่ 2 ถูกพัฒนาขึ้นเพื่อเป็นแนวทางแก้ไขที่สามารถตรวจสอบและคาดการณ์คุณภาพดินได้อย่างแม่นยำ โดยมีรายละเอียดดังนี้

3.1.1 ขอบเขตของต้นแบบนวัตกรรม

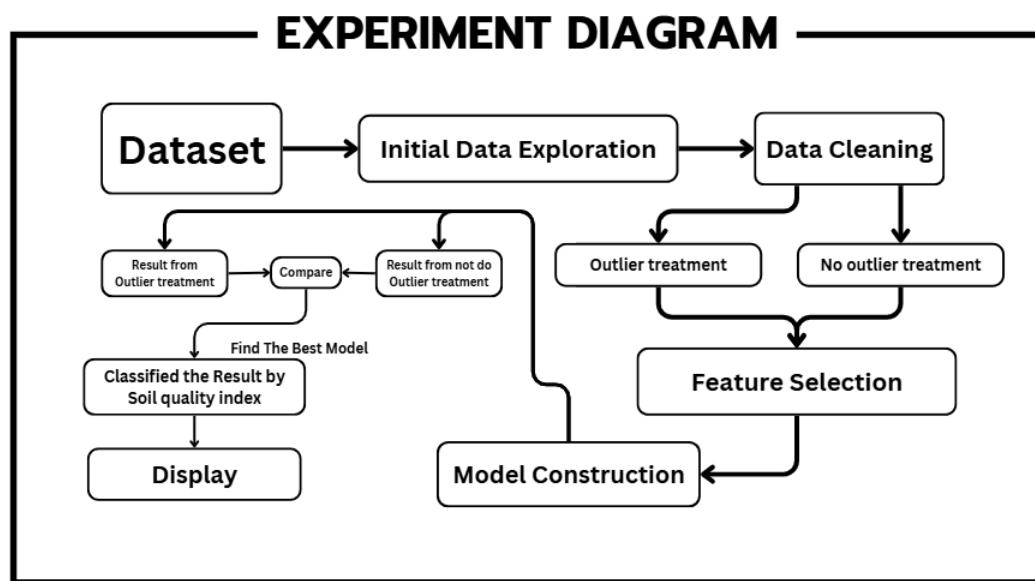
ต้นแบบนวัตกรรมที่พัฒนาขึ้น คือ ระบบวิเคราะห์คุณภาพดินโดยใช้ Machine Learning สำหรับตรวจสอบและคาดการณ์คุณภาพดินในอนาคต โดยต้นแบบนี้มุ่งเน้นเฉพาะส่วนของ Machine Learning Model ในภาระที่ข้อมูลและพยากรณ์โดยใช้ dataset ของทวีปแอฟริกาตอนใต้ ซึ่งขอบเขตการพัฒนาประกอบด้วย

- การออกแบบโมเดลที่สามารถวิเคราะห์คุณภาพดินได้จากการพัฒนาข้อมูล
 - ความสามารถในการนำไฟฟ้าของดิน
 - ความสามารถในการแลกเปลี่ยนไอออนที่มีประจุบวก (cations) ของดิน
 - ค่าความเป็นกรด-ด่าง (pH)
 - ความชื้นในดิน
 - ธาตุต่าง ๆ ที่สำคัญในดิน คือ อลูมิเนียม, ไบرون, แคลเซียม, ทองแดง, เหล็ก, โพแทสเซียม, แมกนีเซียม, แมกนีส, โซเดียม, พอสฟอรัส, ซัลเฟอร์, สังกะสี, ไนโตรเจน, คาร์บอนอินทรีย์ และคาร์บอนอนินทรีย์
 - ส่วนประกอบของดิน คือ ปริมาณดินทราย, ดินร่วน, ดินเหนียว
 - ระดับชั้นของดิน
 - ปริมาณน้ำฝนในพื้นที่
 - อุณหภูมิของชั้นบรรยากาศหนึ่งดิน
 - ความเร็วลม
 - ความสามารถในการอุ้มน้ำและระบายน้ำของดิน

2. การสร้างความแม่นยำในการทำนายด้วยการเปรียบเทียบ 6 ประเภท

โมเดลการเปรียบเทียบ 6 ประเภทได้แก่ Random Forest, Support Vector Regression (SVM), Decision Tree, Gradient Boosting, Extreme Gradient Boosting และ Neural Networks โดยโมเดลใช้ข้อมูลที่ได้จากแหล่งข้อมูลที่มีอยู่ เช่น ISRIC-Africa Soil Profiles Database (AfSP) และ iSDA-Africa

3. การคาดการณ์ของโมเดลจะคาดการณ์ปี ค.ศ. 2012 เท่านั้น จากปีข้อมูลที่มี ค.ศ. 2009 - ค.ศ. 2014
4. ขั้นตอนการทำการทดลองโดยรวม



รูป 7 แผนภาพการทดลองโดยรวม

จาก Dataset ของปี 2009 2010 2011 ที่ได้มานำไปทำ Data Exploration เพื่อทำการกระจายตัวของข้อมูลจากนั้นนำไปทำ Data Cleaning เพื่อจัดการกับข้อมูลที่ขาดหาย จากนั้นทำการทดลองนำข้อมูลที่ผ่านการ Data Cleaning เสร็จมาทดลองแยกเป็นแบบมีการทำ Outlier treatment (การจัดการกับค่าผิดปกติ) แล้วกับแบบไม่มีการทำ Outlier treatment เพื่อทดลองทั้งกรณีที่จัดการกับค่าผิดปกติและไม่จัดการเพื่อหาว่าค่าผิดปกติเป็นค่าที่เครื่องมือวัดทางวิทยาศาสตร์ผิดพลาดหรือไม่โดยจะทำพร้อมกับการ tren โมเดลเพื่อหาความแม่นยำของแต่ละโมเดลร่วมไปด้วย

หลังจากการแบ่งสำเร็จนำไปหา Feature Selection เพื่อหาว่าข้อมูล Feature ใดสัมพันธ์กับ target แต่ละตัวยังไงโดยใช้วิธี MI จากนั้นทำการทดลองนำข้อมูลที่ผ่านการ Outlier treatment แล้ว กับยังไงไม่ผ่านการ

ทำ Outlier treatment มาทดสอบทำนายค่า target ในปี 2012 ผ่าน Model ต่าง ๆ (Random Forest, Support Vector Regression (SVR), Decision Tree, Gradient Boosting, Extreme Gradient Boosting และ Neural Networks) จากนั้นนำค่าที่ทำนายได้จากข้อมูลทั้งสองแบบที่ตัด Outlier กับยังไม่ตัด Outlier มาเปรียบเทียบกันและตรวจสอบโดย Soil quality index เพื่อหาโมเดลที่ทำนายได้ดีที่สุด จากนั้นนำมาแสดงผล

3.1.2 Requirement ของ Prototype

1. ข้อมูลที่ใช้ในการพัฒนาโมเดล

- ชุดข้อมูลคุณภาพดินที่ประกอบด้วยฟีเจอร์ที่สำคัญ ได้แก่
 - ค่าความเป็นกรด-ด่าง (pH)
 - ความชื้นในดิน
 - ความสามารถในการนำไฟฟ้าของดิน
 - ความสามารถในการแลกเปลี่ยน Cations ของดิน
 - ธาตุต่าง ๆ ที่สำคัญในดิน คือ อลูมิเนียม, บอรอน, แคลเซียม, ทองแดง, เหล็ก, โพแทสเซียม, แมกนีเซียม, แมงกานีส, โซเดียม, ฟอสฟอรัส, ซัลเฟอร์, สังกะสี, ไนโตรเจน, คาร์บอน อินทรีย์ และคาร์บอนอนนินทรีย์
 - ส่วนประกอบของดิน คือ ปริมาณดินทราย, ดินร่วน, ดินเหนียว
 - ระดับชั้นของดิน
 - ปริมาณน้ำฝนในพื้นที่
 - อุณหภูมิของชั้นบรรยากาศเหนือดิน
 - ความเร็วลม
 - ความสามารถในการอุ้มน้ำและระบายน้ำของดิน
- ข้อมูลต้องผ่านกระบวนการทำความสะอาด (Data Cleaning) เช่น การจัดการค่าที่หายไป และ การปรับขนาดของข้อมูล (Normalization)

2. เครื่องมือและซอฟต์แวร์

- ภาษาโปรแกรม Python
- ไลบรารีที่ใช้
 - Pandas สำหรับการจัดการข้อมูล
 - NumPy สำหรับการคำนวณทางคณิตศาสตร์
 - Scikit-learn สำหรับการสร้างและวิเคราะห์โมเดล Machine Learning
 - Matplotlib และ seaborn สำหรับการสร้างกราฟและวิเคราะห์ข้อมูล
 - Scipy สำหรับ การสูตรคำนวณความสัมพันธ์ของข้อมูล
 - ไลบรารีเฉพาะของบางโมเดล

3. เกณฑ์การประเมินผล (Evaluation Metrics)

- ค่า Mean Absolute Error (MAE) และ Root Mean Squared Error (RMSE) สำหรับการประเมิน Error ของค่าที่คาดการณ์และค่าจริงที่มี
- ค่า R2 Score เพื่อประเมินว่าจากข้อมูลที่ต้องคาดการณ์นั้น โมเดลใดบាយได้เท่าไหร่และไม่ได้เท่าไหร่

4. การแบ่งข้อมูล (Data Splitting)

แบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่

- ชุดฝึกสอน (Training Set) : 80% ของข้อมูลทั้งหมด
- ชุดทดสอบ (Testing Set) : 20% ของข้อมูลทั้งหมด

โดยวิธีการแบ่งข้อมูล คือ การให้ Training Set เป็นปีที่ต้องการให้โมเดลหาแนวโน้ม โดยในกรณีศึกษานี้จะใช้ปี 2009-2011 เพื่อคาดการณ์ปี 2012

5. ขอบเขตการพัฒนา (Development Scope)

- พัฒนาระบบ Machine Learning ที่สามารถคาดการณ์คุณภาพดินได้
- พื้นที่ที่ทำการทดลองคือทวีปแอฟริกาตอนใต้

3.2 รายละเอียดทางเทคนิค

3.2.1 Data Exploration

1. นำเข้า Data พร้อมกับตรวจสอบความหมายและหน่วยของชุดข้อมูลต่างๆ โดยข้อมูลที่ใช้จะมีที่มาจากการวัดสารเคมีในดินและสภาพแวดล้อมของดินใน ทวีปแอฟริกาให้โดยจะมีรายละเอียดดังนี้

ตาราง 2 ความหมายของ feature

ID	Meaning
EC	ความสามารถการนำไฟฟ้า (dS/m)
M3 Al	ปริมาณ Aluminium ที่ได้จากดินผ่าน Mehlich3 extraction (mg/kg)
M3 B	ปริมาณ Boron ที่ได้จากดินผ่าน Mehlich3 extraction (mg/kg)
M3 Ca	ปริมาณ Calcium ที่ได้จากดินผ่าน Mehlich3 extraction (mg/kg)
M3 Cu	ปริมาณ Copper ที่ได้จากดินผ่าน Mehlich3 extraction (mg/kg)
M3 Fe	ปริมาณ Iron ที่ได้จากดินผ่าน Mehlich3 extraction (mg/kg)
M3 K	ปริมาณ Potassium ที่ได้จากดินผ่าน Mehlich3 extraction (mg/kg)
M3 Mg	ปริมาณ Magnesium ที่ได้จากดินผ่าน Mehlich3 extraction (mg/kg)
M3 Mn	ปริมาณ Manganese ที่ได้จากดินผ่าน Mehlich3 extraction (mg/kg)
M3 Na	ปริมาณ Sodium ที่ได้จากดินผ่าน Mehlich3 extraction (mg/kg)
M3 P	ปริมาณ Phosphorus ที่ได้จากดินผ่าน Mehlich3 extraction (mg/kg)

ID	Meaning
M3 S	ปริมาณ Sulfur ที่ได้จากการดูดซึมผ่าน Mehlich3 extraction (mg/kg)
M3 Zn	ปริมาณ Zinc ที่ได้จากการดูดซึมผ่าน Mehlich3 extraction (mg/kg)
pH	ค่า pH ของดิน
Psa asand	% ทรายในดิน
Psa asilt	% ทรายแบ็งในดิน
Psa aclay	% ดินเหนียวในดิน
Wcvfrsat	ปริมาณสูงสุดที่ดินอิ่มน้ำ (cm ³ water/cm ³ soil)
Wcvfrairdry	ปริมาณช่องอากาศที่มีน้ำไม่อิ่มน้ำในดินแห้งสนิท (cm ³ water/cm ³ soil)
Ksat	ความสามารถการซึมผ่านของน้ำในดิน (cm/hr)
Total nitrogen	% Nitrogen ในดิน
Total carbon	% Carbon ในดิน
Acidified nitrogen	% Nitrogen ที่วัดได้จากการใส่กรดในดิน
Acidified carbon	% Carbon ที่วัดได้จากการใส่กรดในดิน
C % Inorg	% Inorganic carbon ในดิน
C % Org	% Organic carbon ในดิน
ECEC	ความสามารถในการดูดหรือแลกเปลี่ยน Cations (cmolc/ kg soil)
Year	ปีที่ทำการเก็บข้อมูล

ID	Meaning
Month	เดือนที่ทำการเก็บข้อมูล
rfh	ปริมาณน้ำฝนใน 10 วัน (mm)
rfh_avg	ค่าเฉลี่ยปริมาณน้ำฝนรายสาขา (mm)
r1h	ปริมาณน้ำฝนใน 1 เดือนภายในเวลาสະสมจะจบ (mm)
r1h_avg	ค่าเฉลี่ยปริมาณน้ำฝนใน 1 เดือนภายในเวลาสະสมจะจบ (mm)
Temperature	อุณหภูมิเฉลี่ยในแต่ละเดือน (Celcius)
Humidity	% ความชื้นในชั้นบรรยากาศ
Wind Speed	ความเร็วลม

2. ระบุรายละเอียดโดยรวมของการคาดการณ์ที่เราจะทำ

โดยหลักการใช้โมเดล เราจะคาดการณ์สารเคมีในดินทุกตัวของดิน ในปี 2012 ซึ่งข้อมูลที่ต้องการที่จะคาดการณ์มีดังนี้

ตาราง 3 ข้อมูลที่ใช้ในการคาดการณ์

ID	ความหมาย
EC	ความสามารถการนำไฟฟ้า (dS/m)
ECEC	ความสามารถในการแลกเปลี่ยนประจุ Cations ของดิน(dS/m)
M3 Al	ปริมาณ Aluminium (mg/kg)
M3 B	ปริมาณ Boron (mg/kg)
M3 Ca	ปริมาณ Calcium (mg/kg)
M3 Cu	ปริมาณ Copper (mg/kg)
M3 Fe	ปริมาณ Iron (mg/kg)
M3 K	ปริมาณ Potassium (mg/kg)
M3 Mg	ปริมาณ Magnesium (mg/kg)
M3 Mn	ปริมาณ Manganese (mg/kg)
ECEC	ความสามารถในการแลกเปลี่ยนประจุ Cations ของดิน(dS/m)
M3 Al	ปริมาณ Aluminium (mg/kg)
M3 B	ปริมาณ Boron (mg/kg)
M3 Ca	ปริมาณ Calcium (mg/kg)
M3 Na	ปริมาณ Sodium (mg/kg)

ID	ความหมาย
M3 P	ปริมาณ Phosphorus (mg/kg)
M3 S	ปริมาณ Sulfur (mg/kg)
M3 Zn	ปริมาณ Zinc (mg/kg)
pH	ค่า pH ของดิน
Psa asand	% รายในดิน
Psa asilt	% ดินตะกอนในดิน
Psa aclay	% ดินเหนียวในดิน
Wcvfrsat	ปริมาณสูงสุดที่ดินอุ้มน้ำได้
M3 Na	ปริมาณ Sodium (mg/kg)
M3 P	ปริมาณ Phosphorus (mg/kg)
M3 S	ปริมาณ Sulfur (mg/kg)
M3 Zn	ปริมาณ Zinc (mg/kg)
pH	ค่า pH ของดิน
Psa asand	% รายในดิน
Psa asilt	% ดินตะกอนในดิน
Psa aclay	% ดินเหนียวในดิน
Wcvfrsat	ปริมาณสูงสุดที่ดินอุ้มน้ำได้

3. ตรวจสอบประเภทของชุดข้อมูลและประเภทของชุดข้อมูลแต่ละตัวว่าเป็นประเภทอะไรเพื่อการวางแผนการจัดการข้อมูลในอนาคต

#	Column	Non-Null Count	Dtype
0	EC_2012	103 non-null	float64
1	M3 Al_2012	103 non-null	float64
2	M3 B_2012	103 non-null	float64
3	M3 Ca_2012	103 non-null	float64
4	M3 Cu_2012	103 non-null	float64
5	M3 Fe_2012	103 non-null	float64
6	M3 K_2012	103 non-null	float64
7	M3 Mg_2012	103 non-null	float64
8	M3 Mn_2012	103 non-null	float64
9	M3 Na_2012	102 non-null	float64
10	M3 P_2012	103 non-null	float64
11	M3 S_2012	103 non-null	float64
12	M3 Zn_2012	103 non-null	float64
13	PH_2012	103 non-null	float64
14	Psa asand_2012	103 non-null	float64
15	Psa asilt_2012	103 non-null	float64
16	Psa aclay_2012	103 non-null	float64
17	Wcvfrsat_2012	103 non-null	float64
18	Ksat_2012	103 non-null	float64
19	Total nitrogen_2012	103 non-null	float64
20	Total carbon_2012	103 non-null	float64
21	ECEC_2012	103 non-null	float64
22	C % Org_2012	103 non-null	float64
23	C % Inorg_2012	103 non-null	float64

รูป 8 แสดงประเภทข้อมูลของ Target

#	Column	Non-Null Count	Dtype
0	EC_2009	103 non-null	float64
1	M3 Al_2009	103 non-null	float64
2	M3 B_2009	103 non-null	float64
3	M3 Ca_2009	103 non-null	float64
4	M3 Cu_2009	103 non-null	float64
5	M3 Fe_2009	103 non-null	float64
6	M3 K_2009	103 non-null	float64
7	M3 Mg_2009	103 non-null	float64
8	M3 Mn_2009	103 non-null	float64
9	M3 Na_2009	102 non-null	float64
10	M3 P_2009	103 non-null	float64
11	M3 S_2009	103 non-null	float64
12	M3 Zn_2009	103 non-null	float64
13	PH_2009	103 non-null	float64
14	Psa asand_2009	103 non-null	float64
15	Psa asilt_2009	103 non-null	float64
16	Psa aclay_2009	103 non-null	float64
17	Wcvfrsat_2009	102 non-null	float64
18	Wcvfrairdry_2009	102 non-null	float64
19	Ksat_2009	102 non-null	float64
20	Total nitrogen_2009	103 non-null	float64
21	Total carbon_2009	103 non-null	float64
22	Acidified nitrogen_2009	103 non-null	float64
23	Acidified carbon_2009	103 non-null	float64
24	C % Inorg_2009	103 non-null	float64
25	C % Org_2009	103 non-null	float64
26	ECEC_2009	103 non-null	float64
27	Year_2009	103 non-null	int64
28	Month_2009	103 non-null	int64
29	rfh_2009	103 non-null	float64
30	rfh_avg_2009	103 non-null	float64
31	r1h_2009	103 non-null	float64
32	r1h_avg_2009	103 non-null	float64
33	Temperature_2009	103 non-null	float64
34	Humidity_2009	103 non-null	float64
35	Wind Speed_2009	103 non-null	float64

dtypes: float64(34), int64(2)

รูป 9 แสดงประเภทข้อมูลของชุดข้อมูลปี 2009

#	Column	Non-Null Count	Dtype
0	EC_2010	103 non-null	float64
1	M3 Al_2010	103 non-null	float64
2	M3 B_2010	103 non-null	float64
3	M3 Ca_2010	103 non-null	float64
4	M3 Cu_2010	103 non-null	float64
5	M3 Fe_2010	103 non-null	float64
6	M3 K_2010	103 non-null	float64
7	M3 Mg_2010	103 non-null	float64
8	M3 Mn_2010	103 non-null	float64
9	M3 Na_2010	103 non-null	float64
10	M3 P_2010	103 non-null	float64
11	M3 S_2010	103 non-null	float64
12	M3 Zn_2010	103 non-null	float64
13	PH_2010	103 non-null	float64
14	Psa asand_2010	103 non-null	float64
15	Psa asilt_2010	103 non-null	float64
16	Psa aclay_2010	103 non-null	float64
17	Wcvfrsat_2010	103 non-null	float64
18	Wcvfrairdry_2010	103 non-null	float64
19	Ksat_2010	103 non-null	float64
20	Total nitrogen_2010	103 non-null	float64
21	Total carbon_2010	103 non-null	float64
22	Acidified nitrogen_2010	103 non-null	float64
23	Acidified carbon_2010	103 non-null	float64
24	C % Inorg_2010	103 non-null	float64
25	C % Org_2010	103 non-null	float64
26	ECEC_2010	103 non-null	float64
27	Year_2010	103 non-null	int64
28	Month_2010	103 non-null	int64
29	rfh_2010	103 non-null	float64
30	rfh_avg_2010	103 non-null	float64
31	r1h_2010	103 non-null	float64
32	r1h_avg_2010	103 non-null	float64
33	Temperature_2010	103 non-null	float64
34	Humidity_2010	103 non-null	float64
35	Wind Speed_2010	103 non-null	float64

dtypes: float64(34), int64(2)

รูป 10 แสดงปริมาณข้อมูลของชุดข้อมูลปี 2010

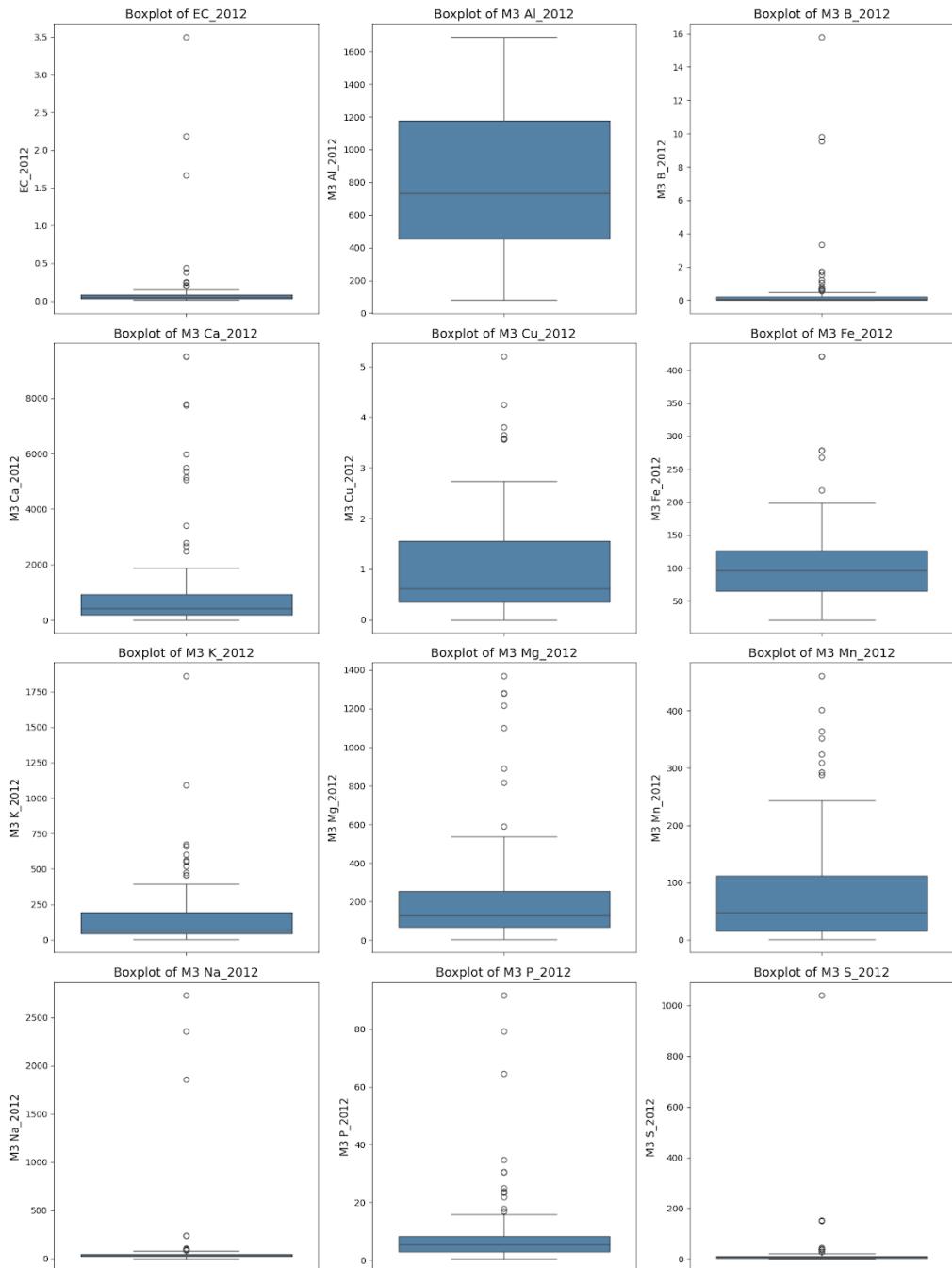
#	Column	Non-Null Count	Dtype
0	EC_2011	103 non-null	float64
1	M3_A1_2011	103 non-null	float64
2	M3_B_2011	103 non-null	float64
3	M3_Ca_2011	103 non-null	float64
4	M3_Cu_2011	103 non-null	float64
5	M3_Fe_2011	103 non-null	float64
6	M3_K_2011	103 non-null	float64
7	M3_Mg_2011	103 non-null	float64
8	M3_Mn_2011	103 non-null	float64
9	M3_Na_2011	101 non-null	float64
10	M3_P_2011	103 non-null	float64
11	M3_S_2011	103 non-null	float64
12	M3_Zn_2011	103 non-null	float64
13	PH_2011	103 non-null	float64
14	Psa_asand_2011	103 non-null	float64
15	Psa_asilt_2011	103 non-null	float64
16	Psa_aclay_2011	103 non-null	float64
17	Wcvfrsat_2011	103 non-null	float64
18	Wcvfrairdry_2011	103 non-null	float64
19	Ksat_2011	103 non-null	float64
20	Total_nitrogen_2011	103 non-null	float64
21	Total_carbon_2011	103 non-null	float64
22	Acidified_nitrogen_2011	103 non-null	float64
23	Acidified_carbon_2011	103 non-null	float64
24	C_%Inorg_2011	103 non-null	float64
25	C_%Org_2011	103 non-null	float64
26	ECEC_2011	103 non-null	float64
27	Year_2011	103 non-null	int64
28	Month_2011	103 non-null	int64
29	rfh_2011	103 non-null	float64
30	rfh_avg_2011	103 non-null	float64
31	r1h_2011	103 non-null	float64
32	r1h_avg_2011	103 non-null	float64
33	Temperature_2011	103 non-null	float64
34	Humidity_2011	103 non-null	float64
35	Wind_Speed_2011	103 non-null	float64

dtypes: float64(34), int64(2)

ຮູບ 11ແສດງປະເກທ໌ຂໍອມູລຂອງໜຸດຂໍອມູລປີ 2011

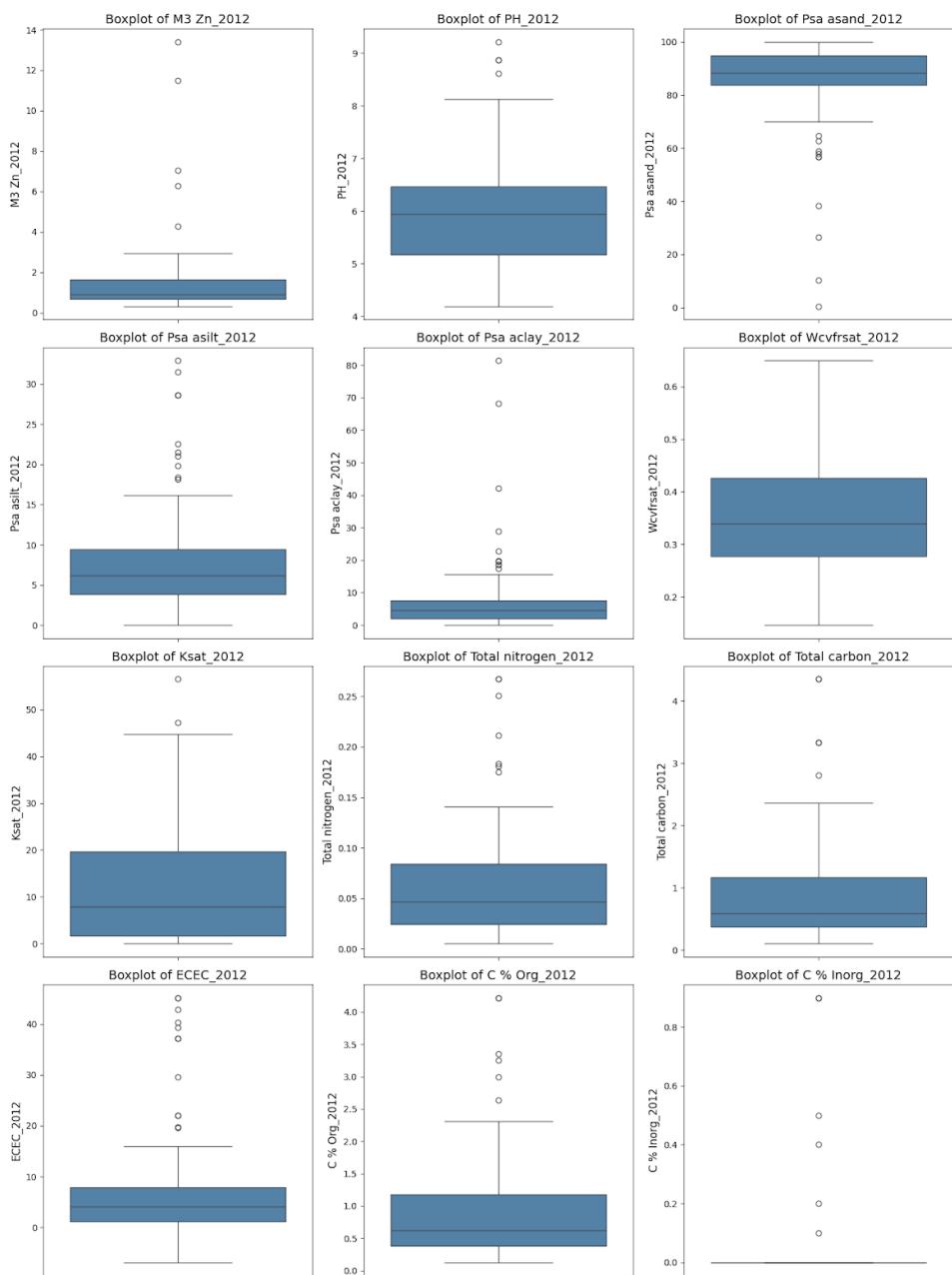
4. วิเคราะห์การกระจายของข้อมูลที่ต้องการคาดการณ์ โดยการ Plot ใช้ Box Plot เนื่องจากข้อมูลเป็นแบบตัวเลขทั้งหมด

Figure 1: Boxplots of First Half of Numeric Columns



รูป 12 การแสดงการกระจายตัวข้อมูลของ target ในปี 2012 (1)

Figure 2: Boxplots of Second Half of Numeric Columns



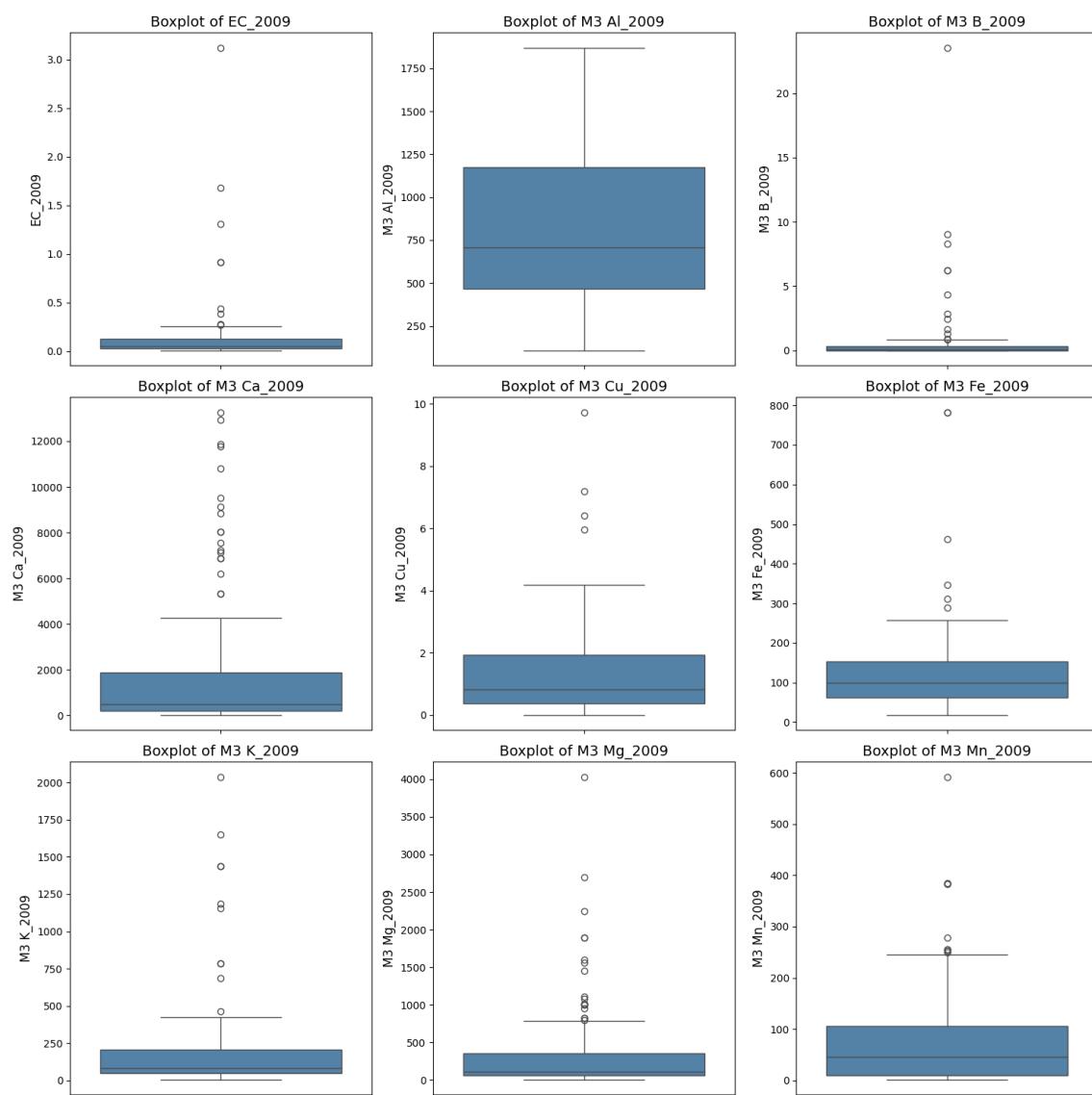
รูป 13 การแสดงการกระจายตัวข้อมูลของ target ในปี 2012 (2)

จากกราฟ Plot สามารถสรุปปี 2012 ได้ว่า

ข้อมูลที่ไม่สมดุลหรือเอียงไปยังฝั่งใดฝั่งหนึ่ง = EC, M3 B, M3 Ca, M3 K, M3 Mg, M3 Na, M3 S, M3 Zn, Total Carbon, Total Nitrogen, Ksat, C % Inorg

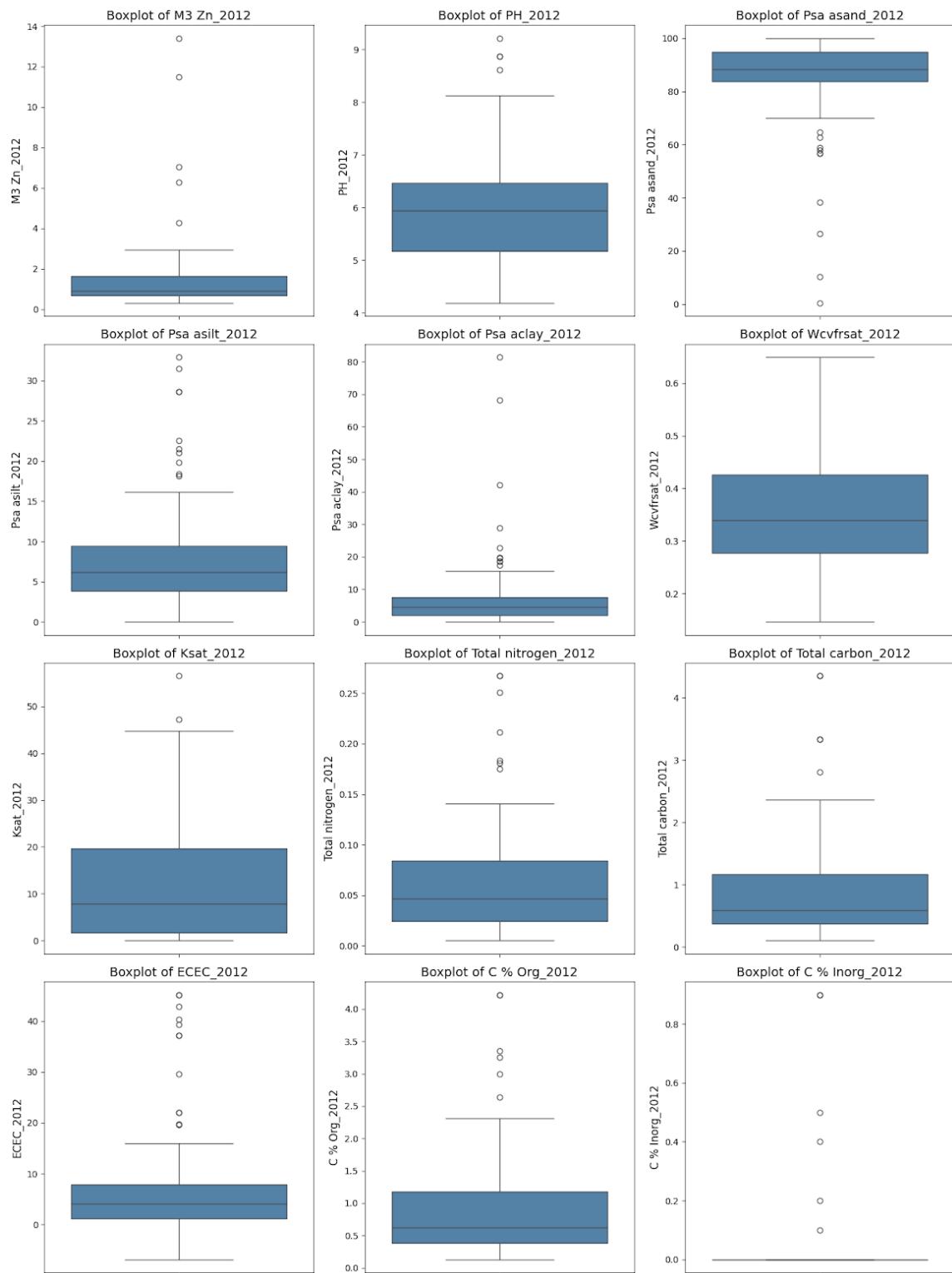
ข้อมูลที่สมดุลและสามารถใช้งานได้เลย = M3 Al, M3 Cu, M3 Mn, Psa aclay, Psa asilt

Figure 1: Boxplots of First Quarter of Numeric Columns (2009)



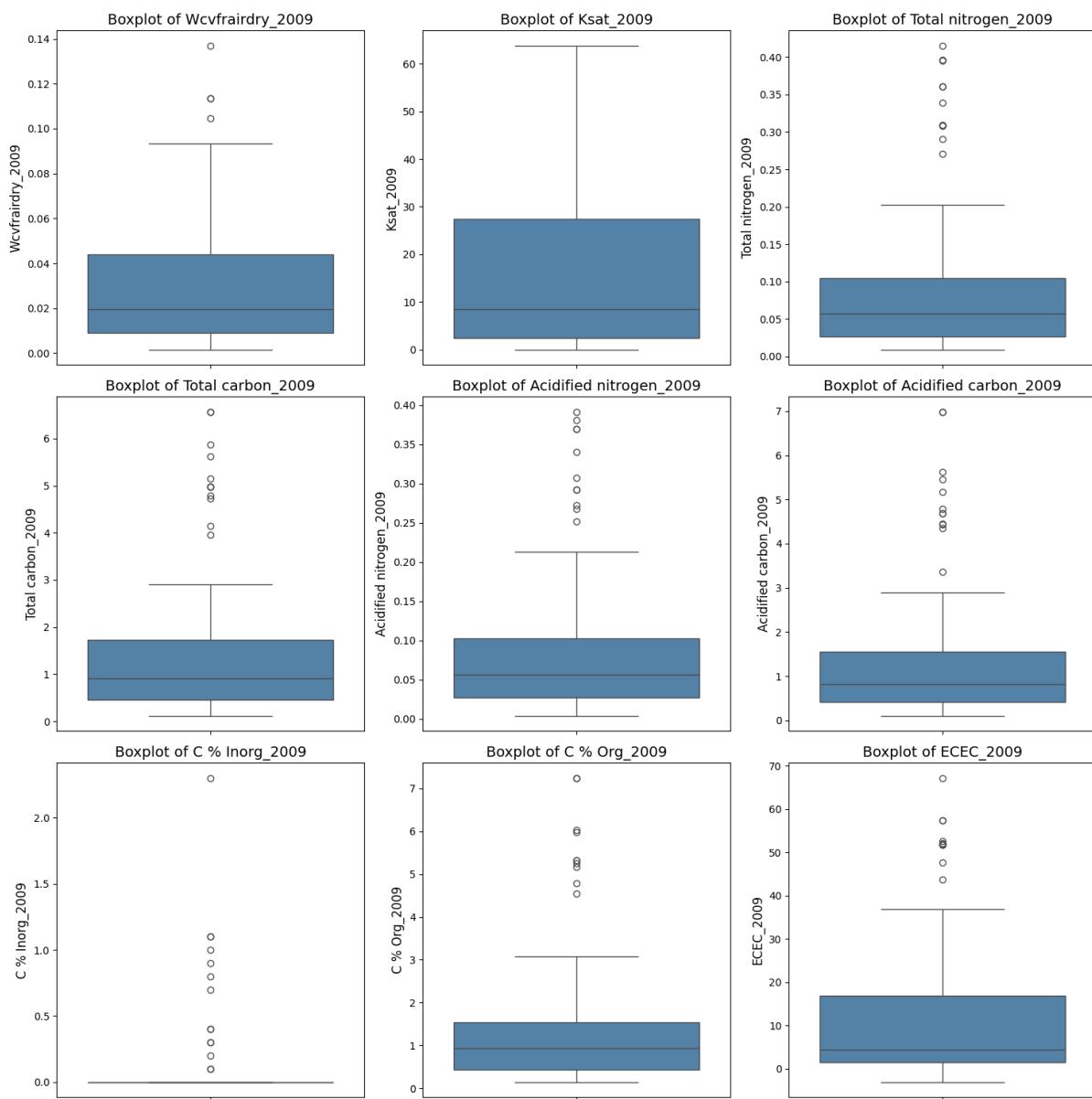
รูป 14 แสดงการกระจายตัวของชุดข้อมูลปี 2009 (1)

Figure 2: Boxplots of Second Half of Numeric Columns



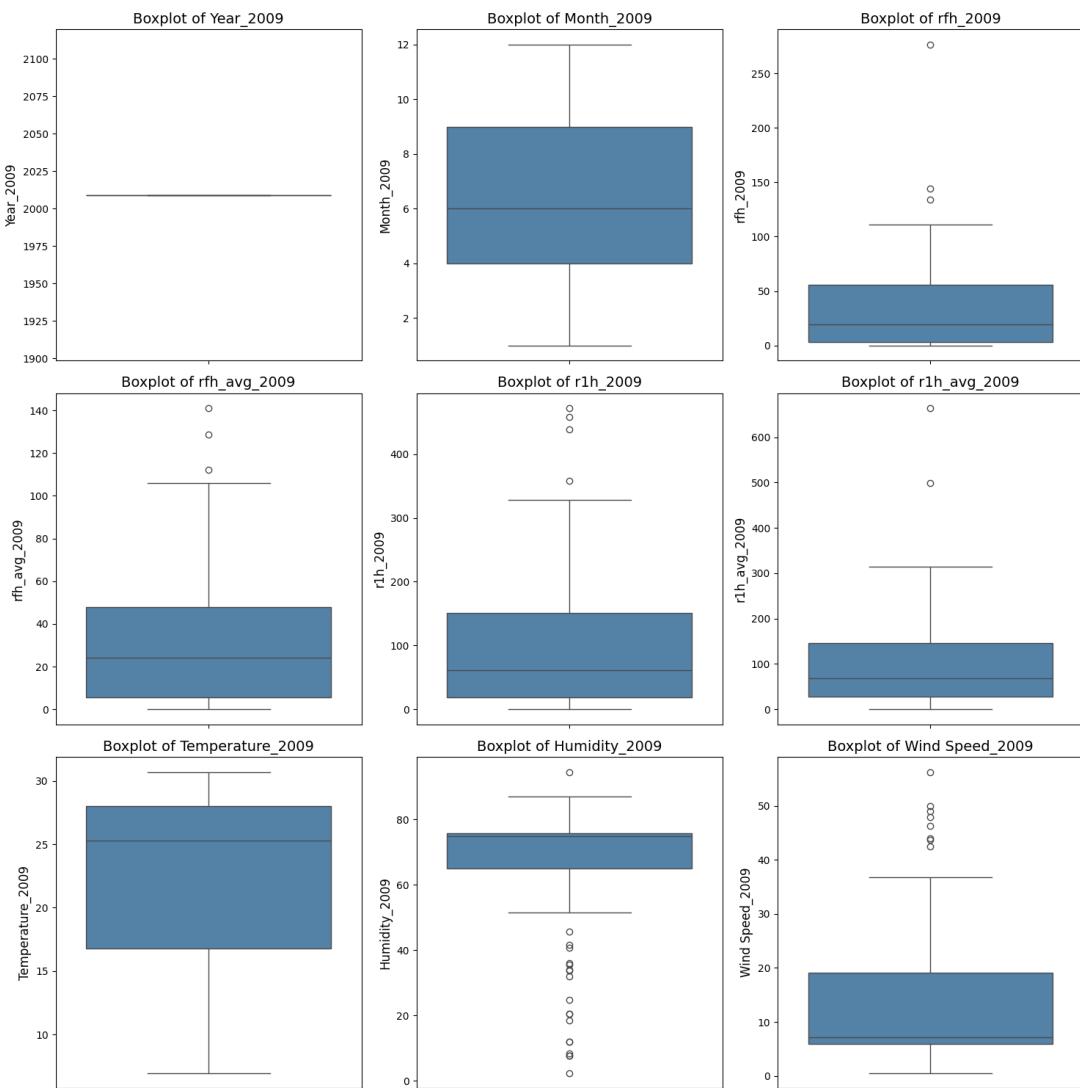
รูป 15 แสดงการกระจายตัวของชุดข้อมูลปี 2009 (2)

Figure 3: Boxplots of Third Quarter of Numeric Columns (2009)



รูป 16 แสดงการกระจายตัวของข้อมูลปี 2009 (3)

Figure 4: Boxplots of Fourth Quarter of Numeric Columns (2009)

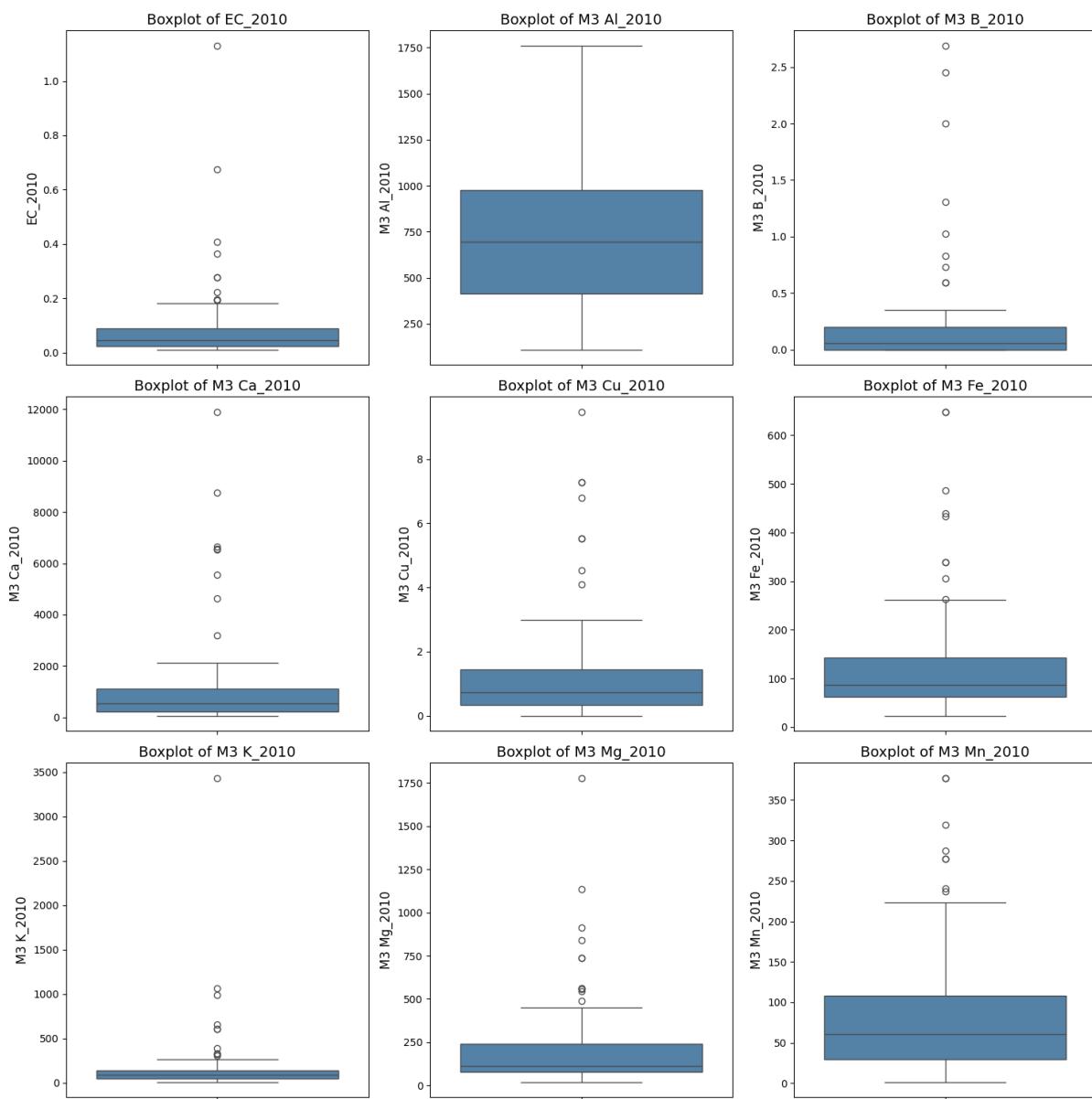


รูป 17 แสดงการกระจายตัวของชุดข้อมูลปี 2009 (4)

จากราฟ Box Plot ด้านบนของปี 2009 สามารถสรุปการกระจายตัวได้ดังนี้

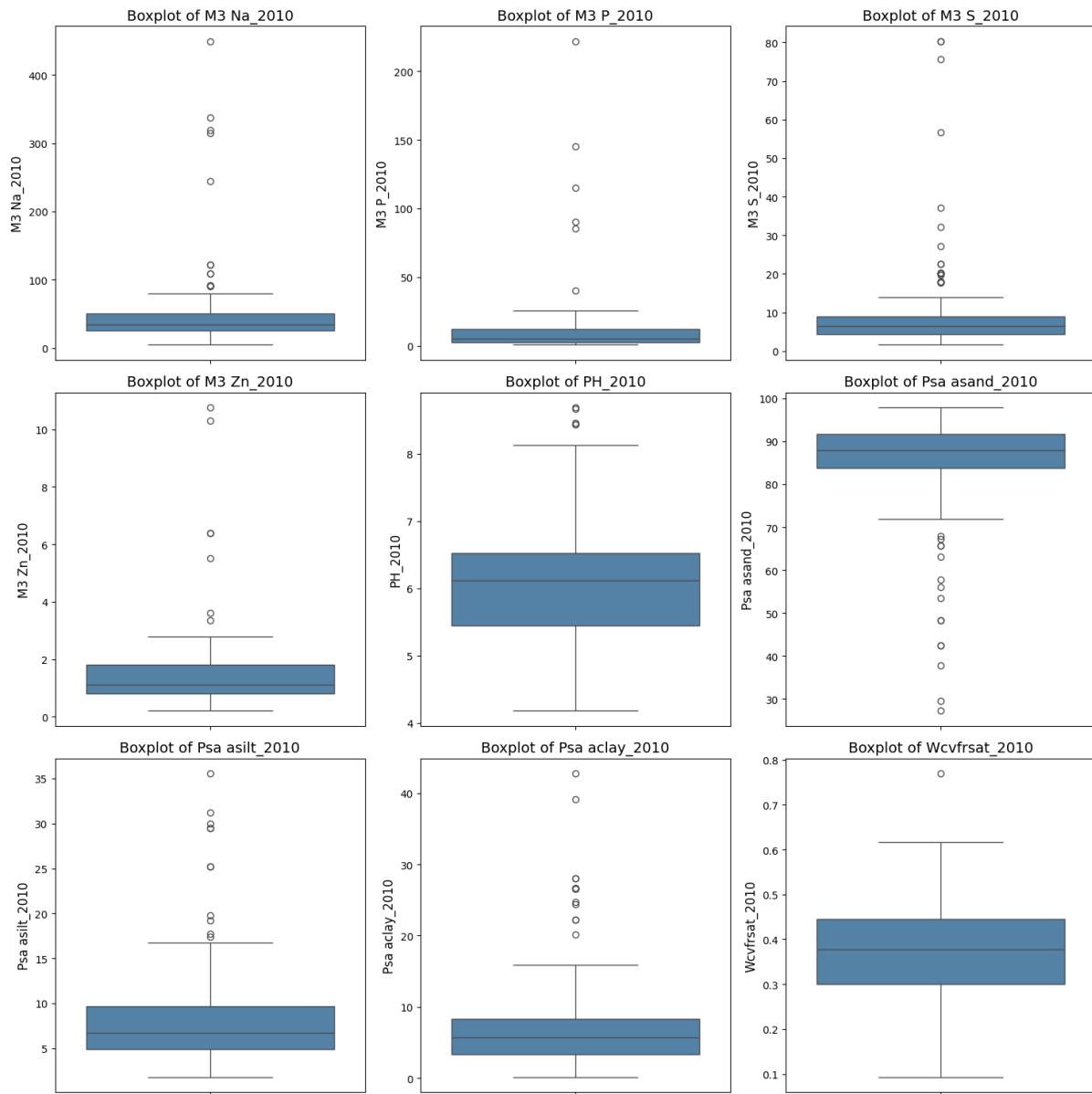
- ข้อมูลที่ไม่สมคุลหรือเอียงไปยังฝั่งใดฝั่งหนึ่ง ได้แก่
EC, M3 B,M3 Ca,M3 Cu,M3 K,M3 Mg,M3 Mn,M3 Na,M3 P,M3 S,M3 Zn,Psa aclay,Total Carbon,Total Nitrogen,Acidified Nitrogen,Acidified Carbon,C % Inorg,C % Org,ECEC,Humidity,Wind Speed
- ข้อมูลที่สมคุลและสามารถใช้งานได้เลย ได้แก่
M3 Al,M3 Fe,PH,Psa asand,Psa asilt,Wcvfrsat,Wcvfrairdry,Ksat, Temperature

Figure 1: Boxplots of First Quarter of Numeric Columns (2010)



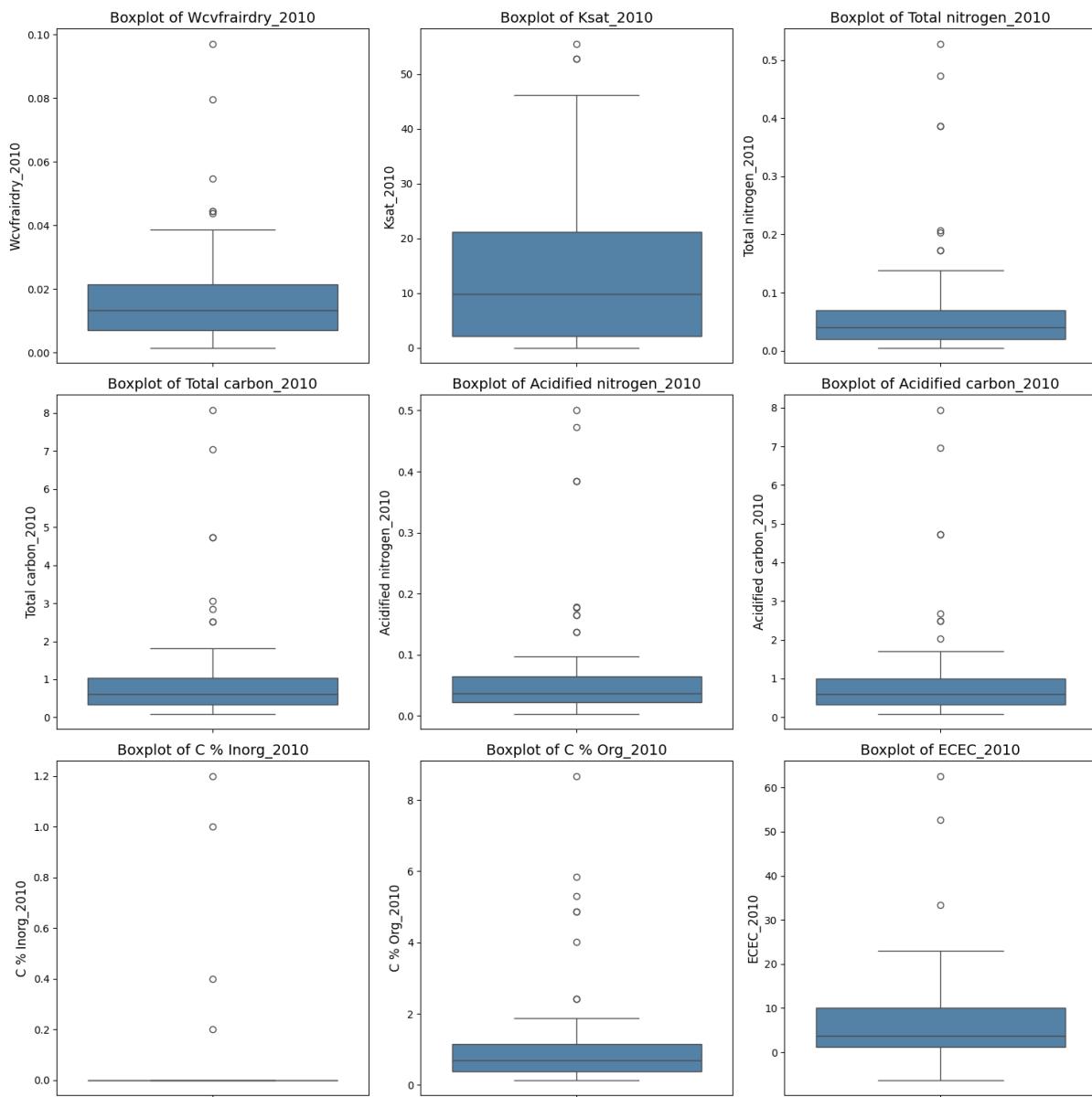
รูป 18 แสดงการกระจายชุดข้อมูลปี 2010 (1)

Figure 2: Boxplots of Second Quarter of Numeric Columns (2010)



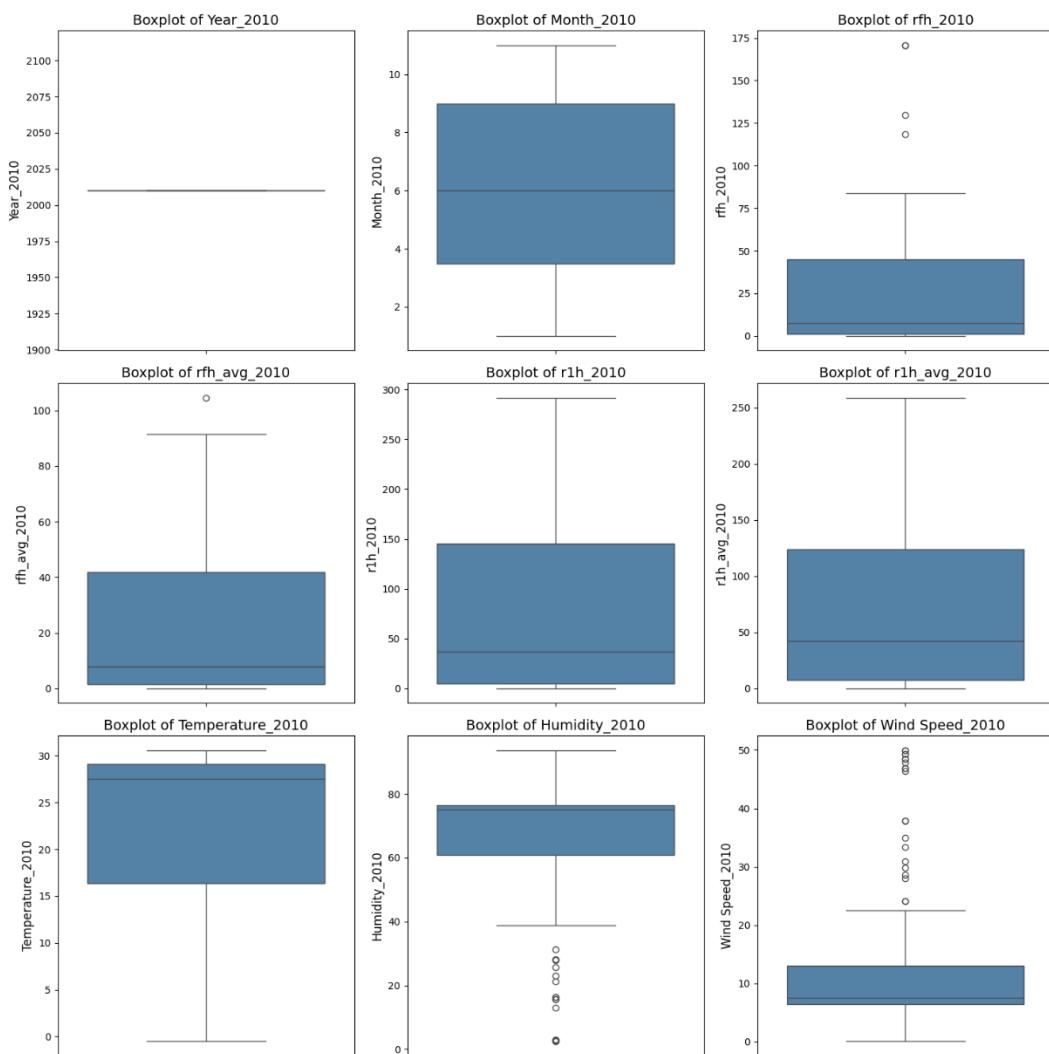
รูป 19 แสดงการกระจายชุดข้อมูลปี 2010 (2)

Figure 3: Boxplots of Third Quarter of Numeric Columns (2010)



รูป 20 แสดงการกระจายชุดข้อมูลปี 2010 (3)

Figure 4: Boxplots of Fourth Quarter of Numeric Columns (2010)

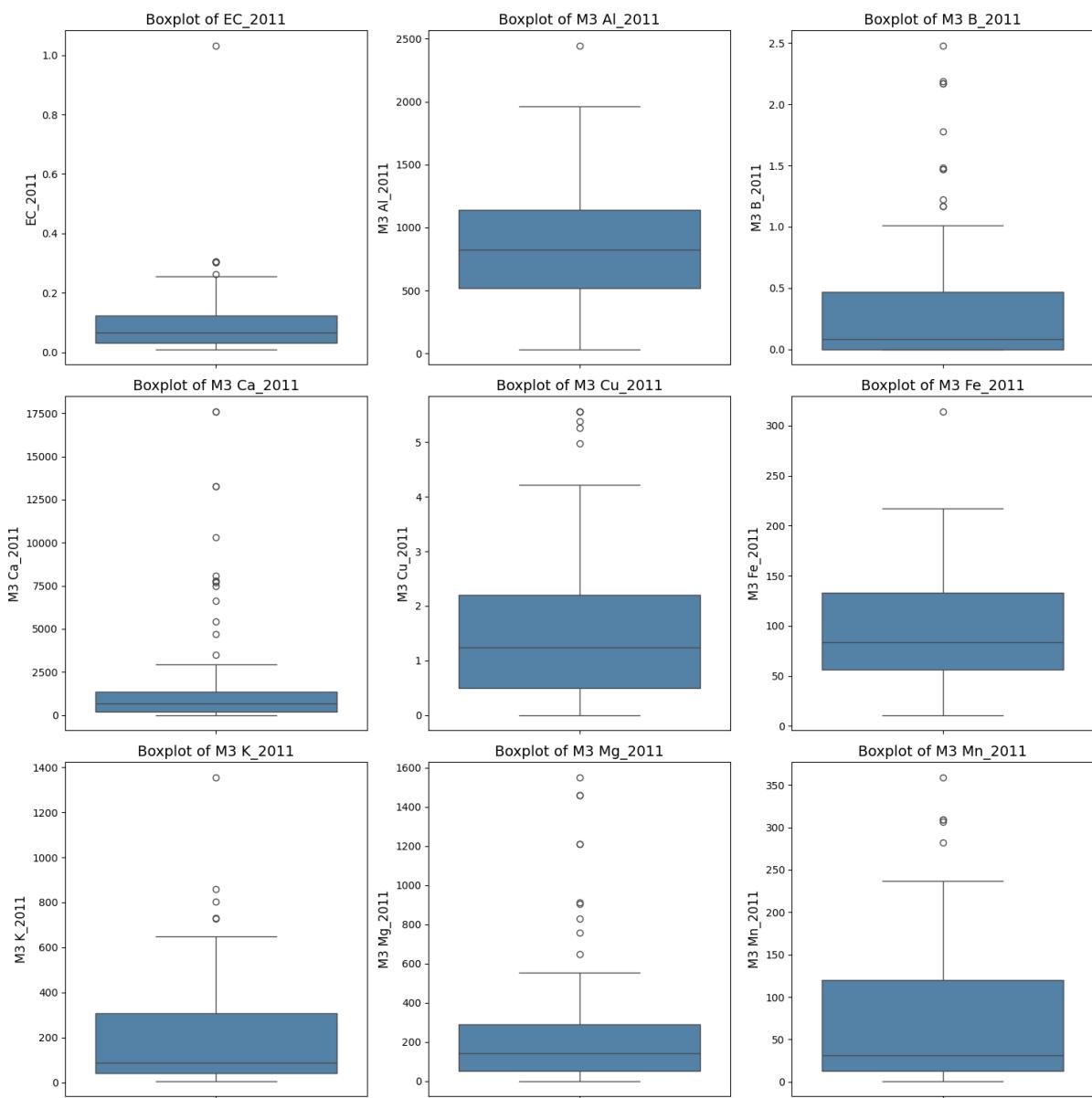


รูป 21 แสดงการกระจายชุดข้อมูลปี 2010 (4)

จากกราฟ Box Plot ด้านบนของปี 2010 สามารถสรุปการกระจายตัวได้ดังนี้

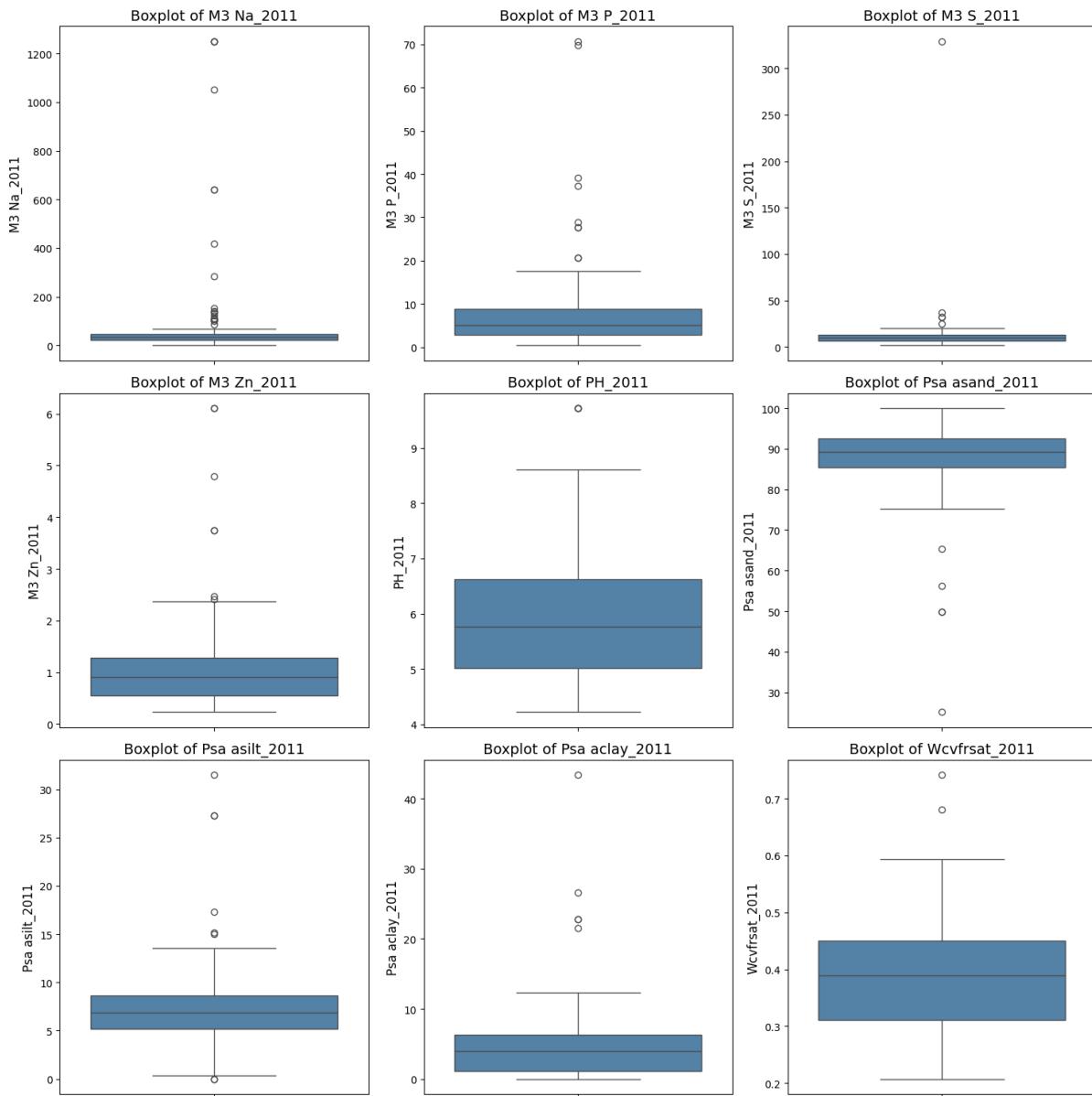
- ข้อมูลที่ไม่สมดุลหรือเอียงไปยังฝั่งใดฝั่งหนึ่ง ได้แก่
EC,M3 B,M3 Ca,M3 Cu,M3 K,M3 Mg,M3 Mn,M3 Na,M3 P,M3 S,M3 Zn,Psa aclay,Total Carbon,Total Nitrogen,Acidified Nitrogen,Acidified Carbon,C % Inorg,C % Org,ECEC,Humidity,Wind Speed
- ข้อมูลที่สมดุลและสามารถใช้งานได้เลย ได้แก่
M3 Al, M3 Fe, PH, Psaasand, Psasilt, Wcvfrsat, Wcvfrairdry, Ksat, Temperature

Figure 1: Boxplots of First Quarter of Numeric Columns (2011)



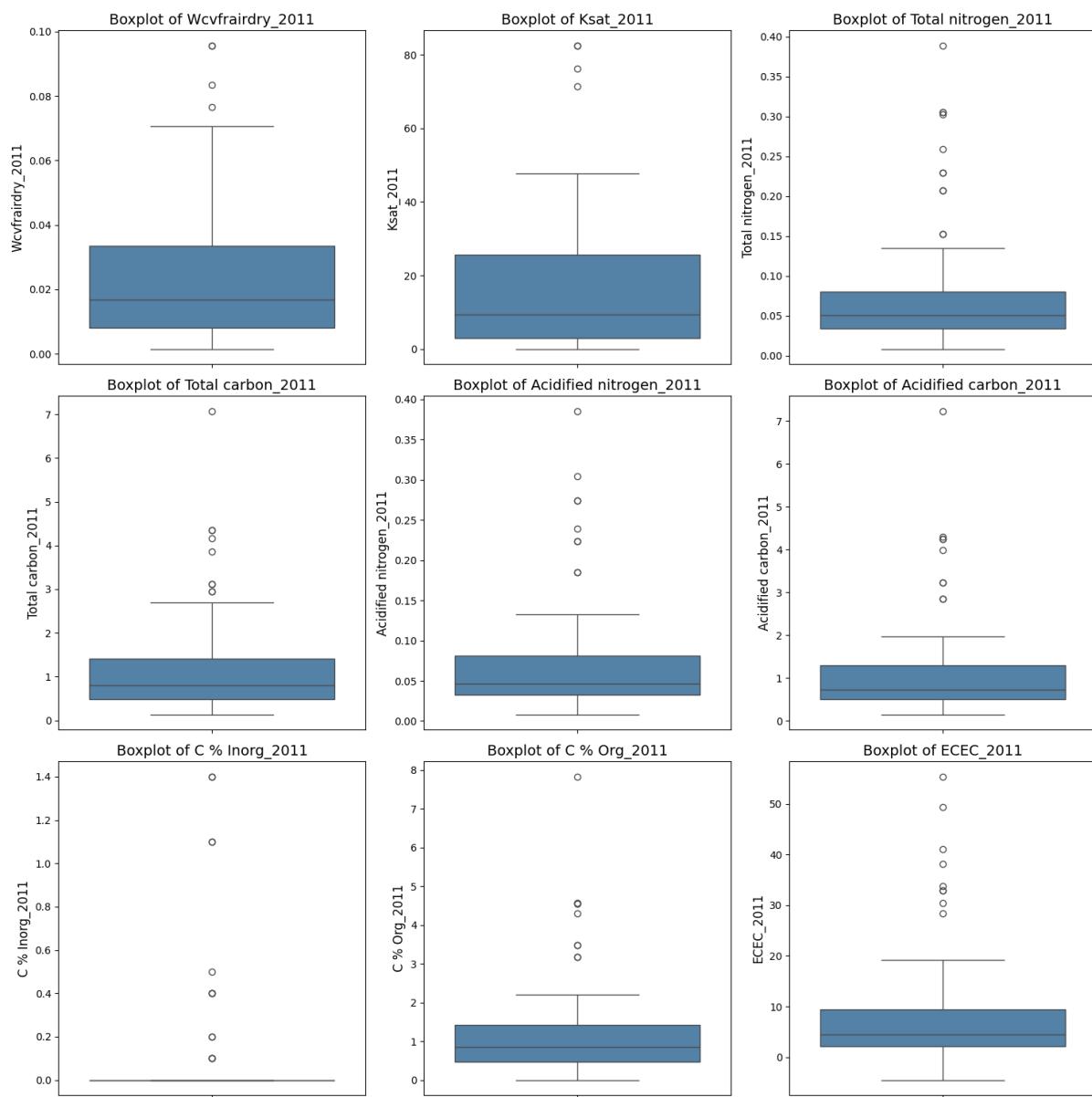
รูป 22 แสดงการกระจายชุดข้อมูลปี 2011(1)

Figure 2: Boxplots of Second Quarter of Numeric Columns (2011)



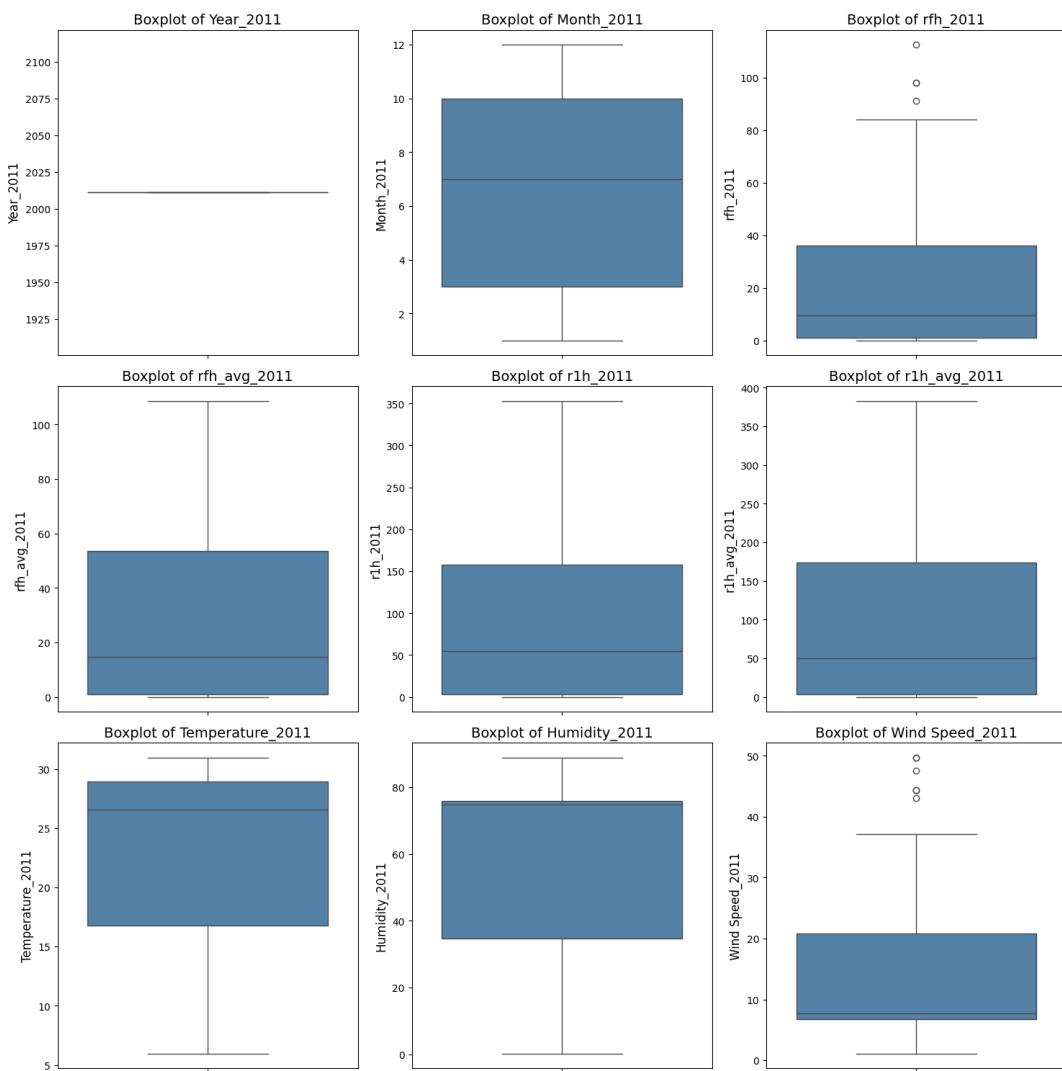
รูป 23 แสดงการกระจำชุดข้อมูลปี 2011(2)

Figure 3: Boxplots of Third Quarter of Numeric Columns (2011)



รูป 24 แสดงการกระจายชุดข้อมูลปี 2011(3)

Figure 4: Boxplots of Fourth Quarter of Numeric Columns (2011)



รูป 25 แสดงการกระจายข้อมูลปี 2011(4)

จากราฟ Box Plot ด้านบนของปี 2011 สามารถสรุปการกระจายตัวได้ดังนี้

- ข้อมูลที่ไม่สมดุลหรือเอียงไปยังฝั่งใดฝั่งหนึ่ง ได้แก่
EC,M3 B,M3 Ca,M3 Cu,M3 K,M3 Mg,M3 Mn,M3 Na,M3 P,M3 S,M3 Zn,Psa aclay,Total Carbon,Total Nitrogen,Acidified Nitrogen,Acidified Carbon,C % Inorg,C % Org,ECEC,Wind Speed
- ข้อมูลที่สมดุลและสามารถใช้งานได้เลย ได้แก่
M3 Al,M3 Fe,PH,Psa asand,Psa asilt,Wcvfrsat,Wcvfrairdry,Ksat,Temperature,Humidity

3.2.2 Data Preprocessing

1. Missing Value

ใช้ `.isna().sum()` เพื่อหาผลรวมของค่าที่หายไปในแต่ละชุดข้อมูลของทุกปี

EC_2009	0	EC_2010	0	EC_2011	0	EC_2012	0
M3 Al_2009	0	M3 Al_2010	0	M3 Al_2011	0	M3 Al_2012	0
M3 B_2009	0	M3 B_2010	0	M3 B_2011	0	M3 B_2012	0
M3 Ca_2009	0	M3 Ca_2010	0	M3 Ca_2011	0	M3 Ca_2012	0
M3 Cu_2009	0	M3 Cu_2010	0	M3 Cu_2011	0	M3 Cu_2012	0
M3 Fe_2009	0	M3 Fe_2010	0	M3 Fe_2011	0	M3 Fe_2012	0
M3 K_2009	0	M3 K_2010	0	M3 K_2011	0	M3 K_2012	0
M3 Mg_2009	0	M3 Mg_2010	0	M3 Mg_2011	0	M3 Mg_2012	0
M3 Mn_2009	0	M3 Mn_2010	0	M3 Mn_2011	0	M3 Mn_2012	0
M3 Na_2009	1	M3 Na_2010	0	M3 Na_2011	2	M3 Na_2012	1
M3 P_2009	0	M3 P_2010	0	M3 P_2011	0	M3 P_2012	0
M3 S_2009	0	M3 S_2010	0	M3 S_2011	0	M3 S_2012	0
M3 Zn_2009	0	M3 Zn_2010	0	M3 Zn_2011	0	M3 Zn_2012	0
PH_2009	0	PH_2010	0	PH_2011	0	PH_2012	0
Psa asand_2009	0	Psa asand_2010	0	Psa asand_2011	0	Psa asand_2012	0
Psa asilt_2009	0	Psa asilt_2010	0	Psa asilt_2011	0	Psa asilt_2012	0
Psa aclay_2009	0	Psa aclay_2010	0	Psa aclay_2011	0	Psa aclay_2012	0
Wcvfrsat_2009	1	Wcvfrsat_2010	0	Wcvfrsat_2011	0	Wcvfrsat_2012	0
Wcvfrairdry_2009	1	Wcvfrairdry_2010	0	Wcvfrairdry_2011	0	Wcvfrairdry_2012	0
Ksat_2009	1	Ksat_2010	0	Ksat_2011	0	Ksat_2012	0
Total nitrogen_2009	0	Total nitrogen_2010	0	Total nitrogen_2011	0	Total nitrogen_2012	0
Total carbon_2009	0	Total carbon_2010	0	Total carbon_2011	0	Total carbon_2012	0
Acidified nitrogen_2009	0	Acidified nitrogen_2010	0	Acidified nitrogen_2011	0	Acidified nitrogen_2012	0
Acidified carbon_2009	0	Acidified carbon_2010	0	Acidified carbon_2011	0	Acidified carbon_2012	0
C % Inorg_2009	0	C % Inorg_2010	0	C % Inorg_2011	0	C % Inorg_2012	0
C % Org_2009	0	C % Org_2010	0	C % Org_2011	0	C % Org_2012	0
ECEC_2009	0	ECEC_2010	0	ECEC_2011	0	ECEC_2012	0

รูป 26 missing value ทุกปี

การขาดหายของข้อมูลจะแก้โดยการแทนที่ค่าที่ขาดหายโดยการใส่ค่าเฉลี่ยของชุดข้อมูลนั้นลงไป

โดยใช้ Function SimpleImputer(strategy='mean')

EC_2009	0	EC_2010	0	EC_2011	0	EC_2012	0
M3 Al_2009	0	M3 Al_2010	0	M3 Al_2011	0	M3 Al_2012	0
M3 B_2009	0	M3 B_2010	0	M3 B_2011	0	M3 B_2012	0
M3 Ca_2009	0	M3 Ca_2010	0	M3 Ca_2011	0	M3 Ca_2012	0
M3 Cu_2009	0	M3 Cu_2010	0	M3 Cu_2011	0	M3 Cu_2012	0
M3 Fe_2009	0	M3 Fe_2010	0	M3 Fe_2011	0	M3 Fe_2012	0
M3 K_2009	0	M3 K_2010	0	M3 K_2011	0	M3 K_2012	0
M3 Mg_2009	0	M3 Mg_2010	0	M3 Mg_2011	0	M3 Mg_2012	0
M3 Mn_2009	0	M3 Mn_2010	0	M3 Mn_2011	0	M3 Mn_2012	0
M3 Na_2009	0	M3 Na_2010	0	M3 Na_2011	0	M3 Na_2012	0
M3 P_2009	0	M3 P_2010	0	M3 P_2011	0	M3 P_2012	0
M3 S_2009	0	M3 S_2010	0	M3 S_2011	0	M3 S_2012	0
M3 Zn_2009	0	M3 Zn_2010	0	M3 Zn_2011	0	M3 Zn_2012	0
PH_2009	0	PH_2010	0	PH_2011	0	PH_2012	0
Psa asand_2009	0	Psa asand_2010	0	Psa asand_2011	0	Psa asand_2012	0
Psa asilt_2009	0	Psa asilt_2010	0	Psa asilt_2011	0	Psa asilt_2012	0
Psa aclay_2009	0	Psa aclay_2010	0	Psa aclay_2011	0	Psa aclay_2012	0
Wcvfrsat_2009	0	Wcvfrsat_2010	0	Wcvfrsat_2011	0	Wcvfrsat_2012	0
Wcvfrairdry_2009	0	Wcvfrairdry_2010	0	Wcvfrairdry_2011	0	Wcvfrairdry_2012	0
Ksat_2009	0	Ksat_2010	0	Ksat_2011	0	Ksat_2012	0
Total nitrogen_2009	0	Total nitrogen_2010	0	Total nitrogen_2011	0	Total nitrogen_2012	0
Total carbon_2009	0	Total carbon_2010	0	Total carbon_2011	0	Total carbon_2012	0
Acidified nitrogen_2009	0	Acidified nitrogen_2010	0	Acidified nitrogen_2011	0	Acidified nitrogen_2012	0
Acidified carbon_2009	0	Acidified carbon_2010	0	Acidified carbon_2011	0	Acidified carbon_2012	0
C % Inorg_2009	0	C % Inorg_2010	0	C % Inorg_2011	0	C % Inorg_2012	0
C % Org_2009	0	C % Org_2010	0	C % Org_2011	0	C % Org_2012	0
ECEC_2009	0	ECEC_2010	0	ECEC_2011	0	ECEC_2012	0

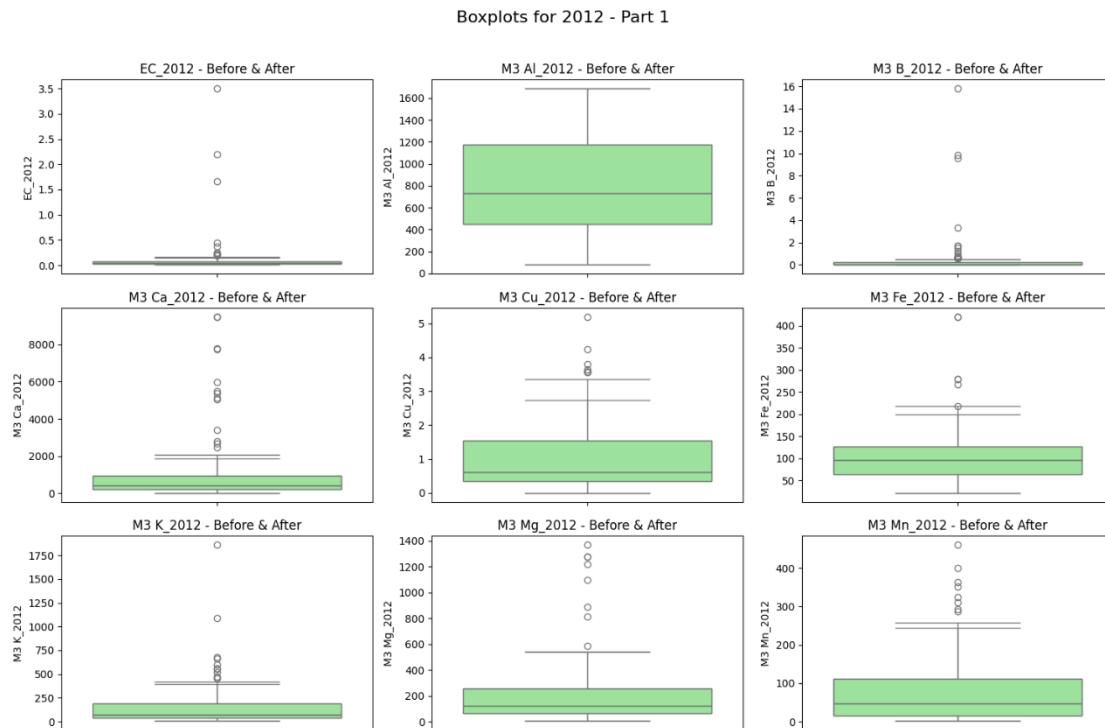
รูป 27 missing value ของแต่ละชุดข้อมูลหลังจากใส่ค่าเฉลี่ยแต่ละชุดข้อมูล

2. Outlier Treatment

เนื่องจากการทางกลุ่มล้มเหลวในการยืนขอข้อมูลความแม่นยำของอุปกรณ์ทางวิทยาศาสตร์เพื่อเหตุผลบางอย่างผู้ทำการทดลองทำให้ทางกลุ่มต้องตั้งสมมติแยกการจัดการค่าผิดปกติเป็น 2 กรณีคือ

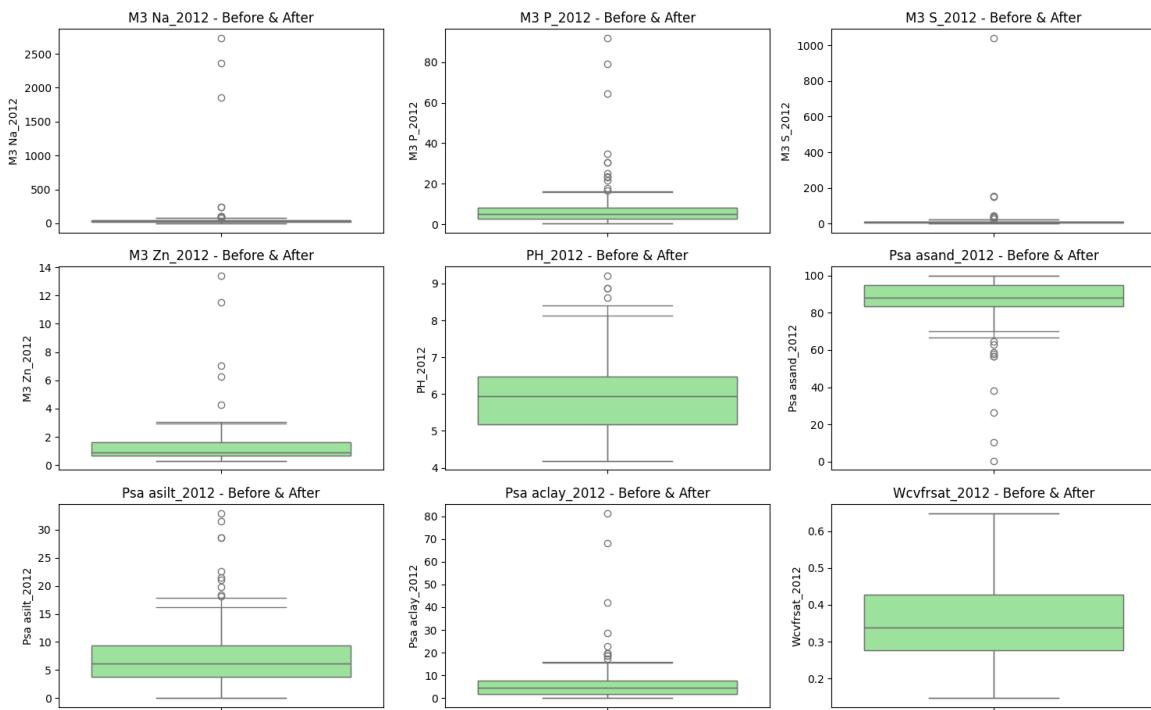
- เครื่องมือวัดทางวิทยาศาสตร์เกิดความผิดพลาดทำให้มีค่าที่ได้มีความผิดปกติโดยมีรีสิแก๊กี้ดังนี้ใช้ Function skew() ที่วัดค่าความผิดปกติของการกระจายตัวข้อมูลทุกตัวซึ่งค่าที่ได้มี 3 กรณีดังนี้
 - a. $\text{skew} > 0$ ชุดข้อมูลมีปริมาณของค่าข้อมูลที่สูงเยอะกว่าปริมาณของค่าข้อมูลที่ต่ำ
 - b. $\text{skew} = 0$ ชุดข้อมูลมีการกระจายตัวที่สมดุลสามารถนำไปใช้ได้เลย
 - c. $\text{skew} < 0$ ชุดข้อมูลมีปริมาณของค่าข้อมูลที่สูงน้อยกว่าปริมาณของค่าข้อมูลที่ต่ำ

หลังจากที่รู้ว่าชุดข้อมูลใดที่มีการกระจายตัวผิดปกติโปรแกรมก็จะใช้ IQR หรือการตั้งขอบเขตของช่วงข้อมูลเพื่อป้องกันไม่ให้ข้อมูลมีค่าเกินกว่าช่วงที่กำหนด เพื่อแก้ไขปัญหาข้อมูลมีการกระจายตัวผิดปกติซึ่งผลลัพธ์ที่ได้มีดังนี้



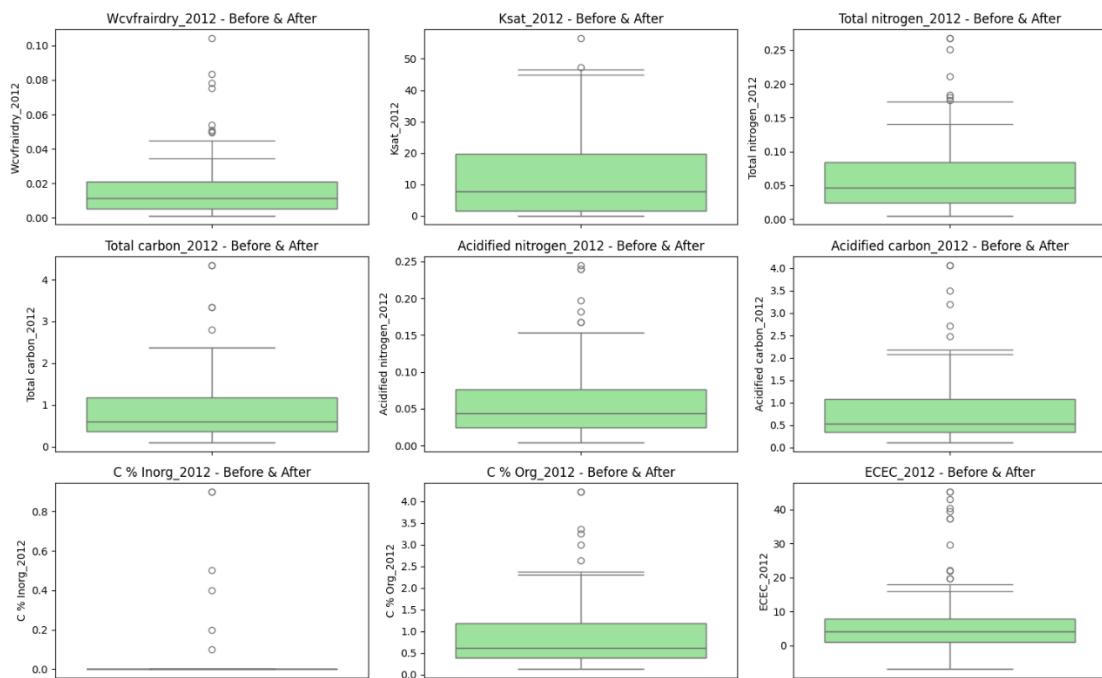
รูป 28 การกระจายตัวของข้อมูล ปี 2012 หลังการจัดการข้อมูลผิดปกติ (1)

Boxplots for 2012 - Part 2



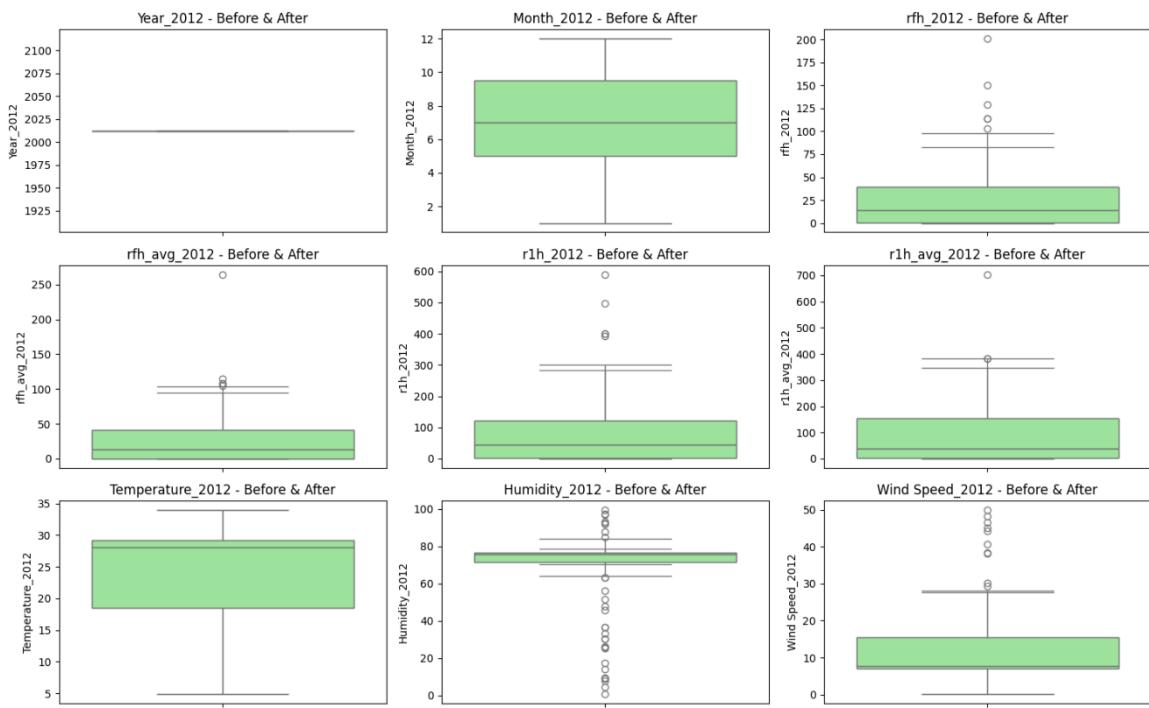
รูป 29 การกระจายตัวของข้อมูล ปี 2012 หลังการจัดการข้อมูลผิดปกติ (2)

Boxplots for 2012 - Part 3



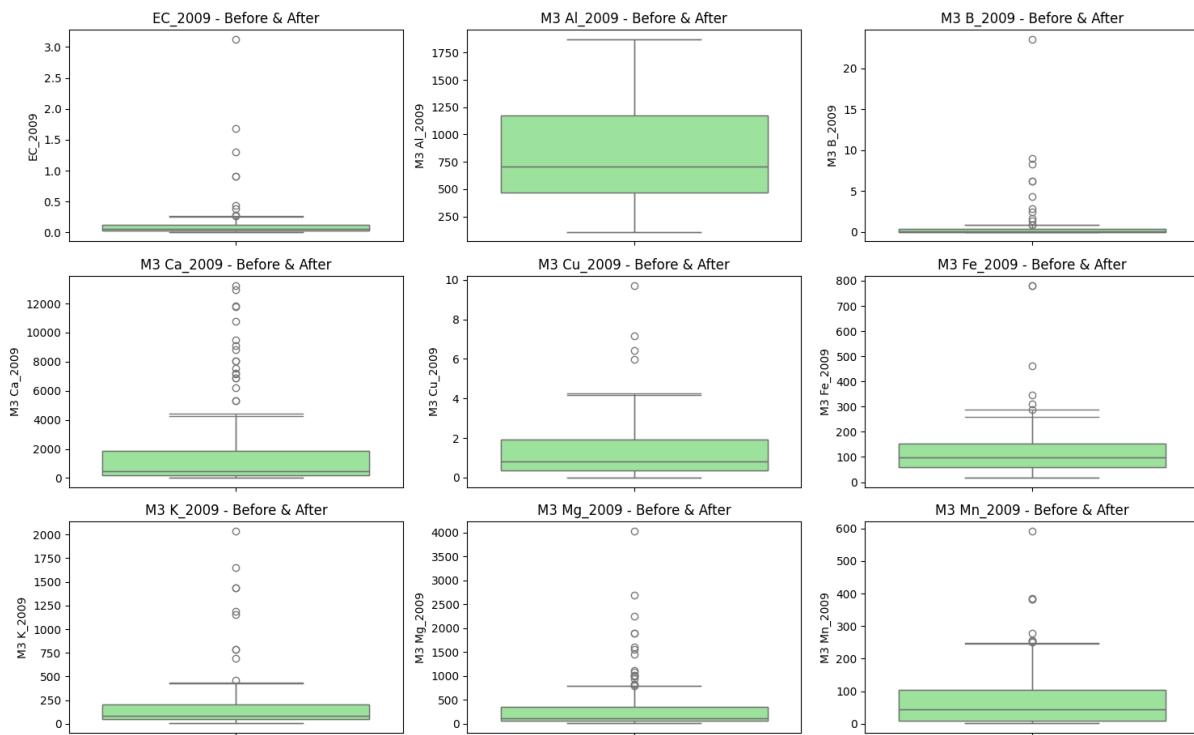
รูป 30 การกระจายตัวของข้อมูล ปี 2012 หลังการจัดการข้อมูลผิดปกติ (3)

Boxplots for 2012 - Part 4



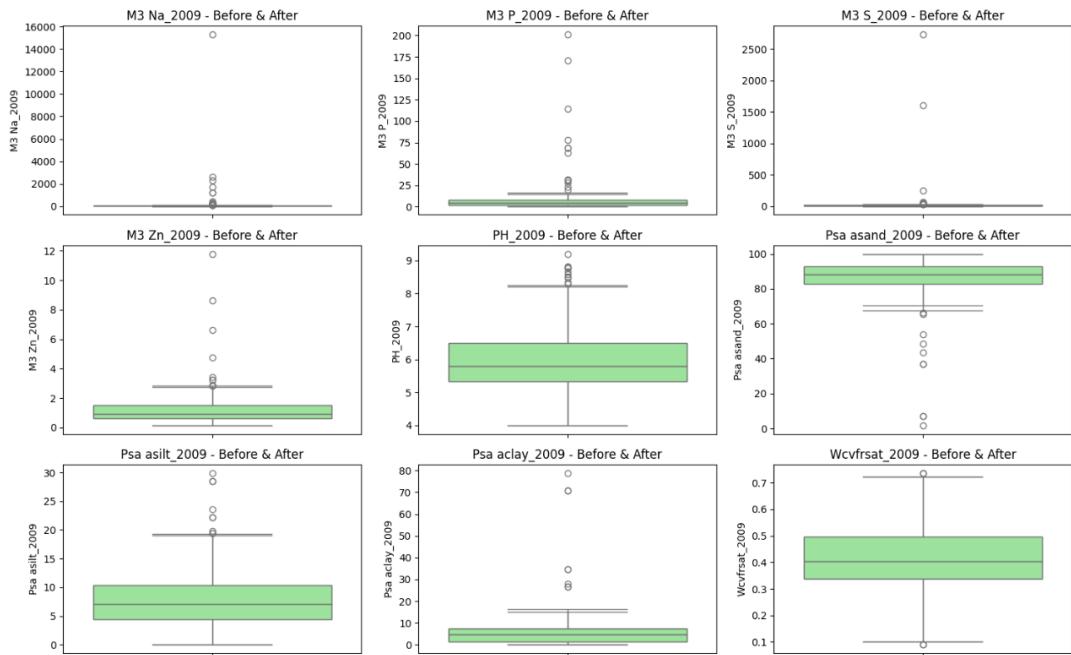
รูป 31 การกระจายตัวของข้อมูล ปี 2012 หลังการจัดการข้อมูลพิเศษ (4)

Boxplots for 2009 - Part 1



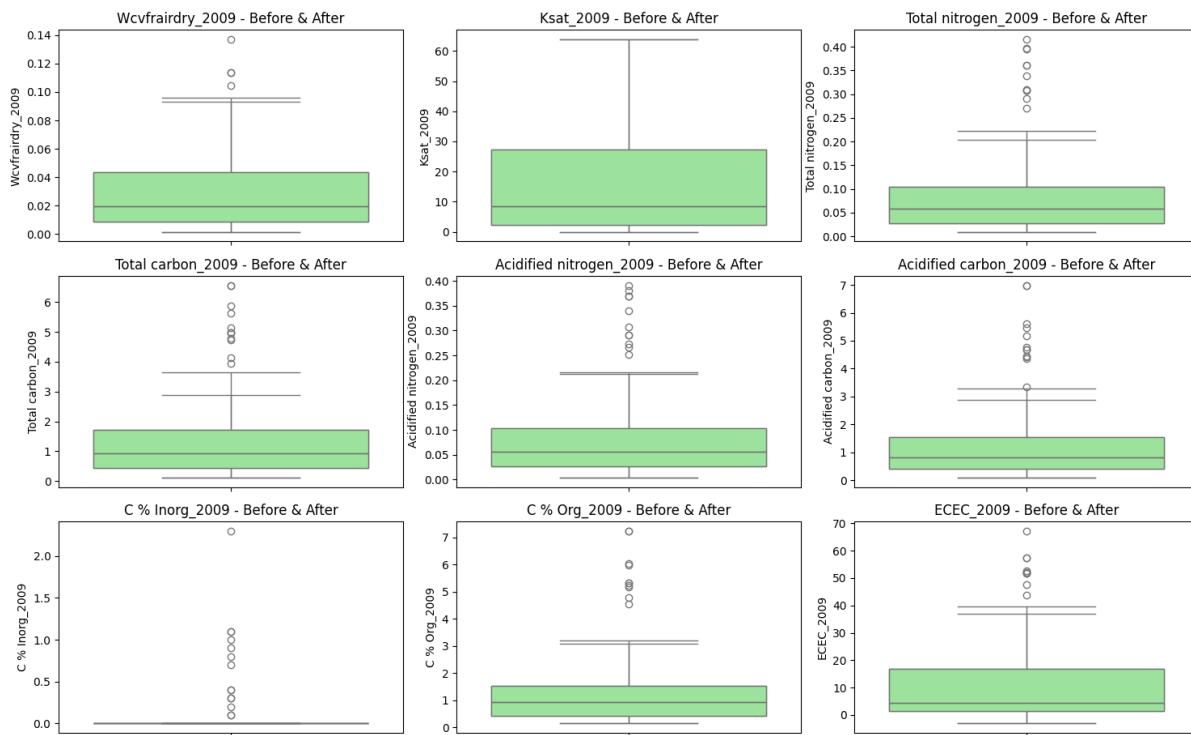
รูป 32 การกระจายตัวของข้อมูล ปี 2009 หลังการจัดการข้อมูลผิดปกติ (1)

Boxplots for 2009 - Part 2



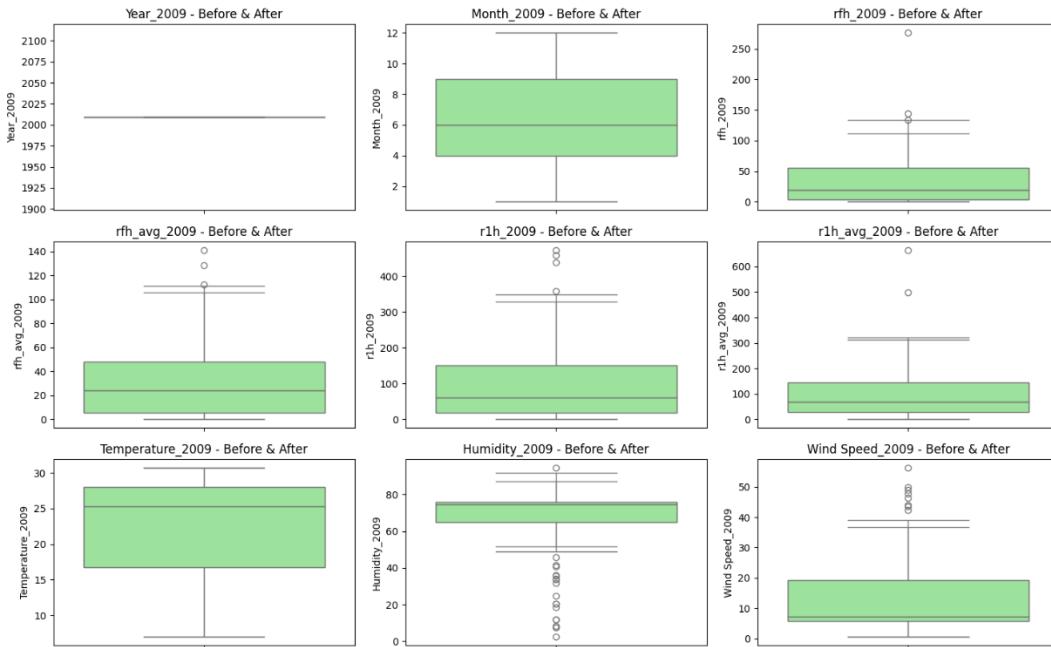
รูป 33 การกระจายตัวของข้อมูล ปี 2009 หลังการจัดการข้อมูลผิดปกติ (2)

Boxplots for 2009 - Part 3



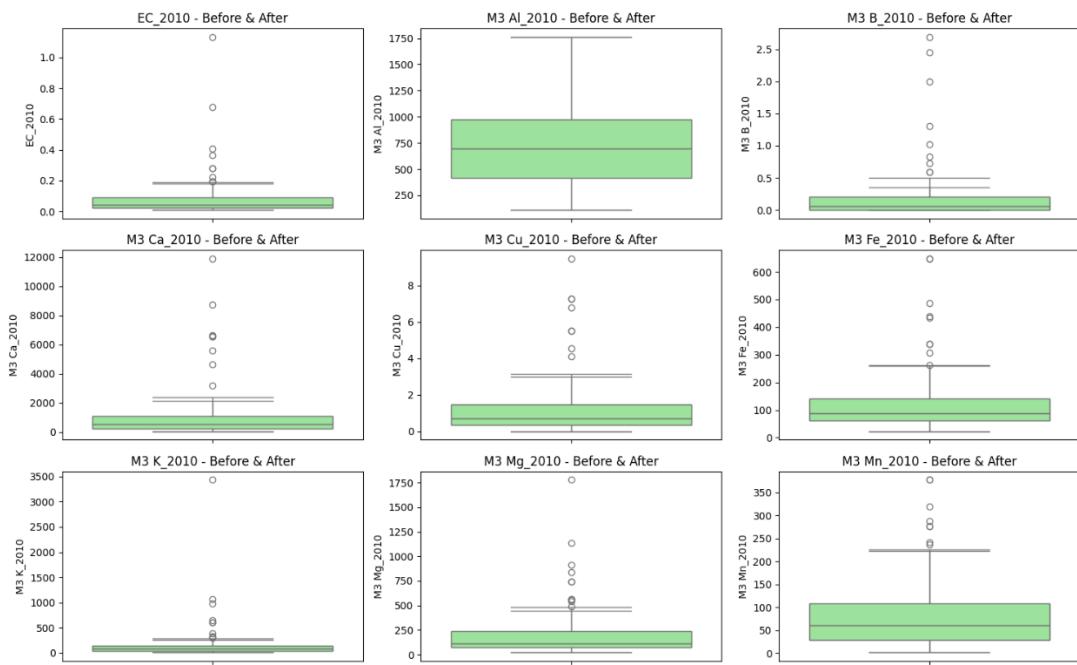
รูป 34 การกระจายตัวของข้อมูล ปี 2009 หลังการจัดการข้อมูลผิดปกติ (3)

Boxplots for 2009 - Part 4



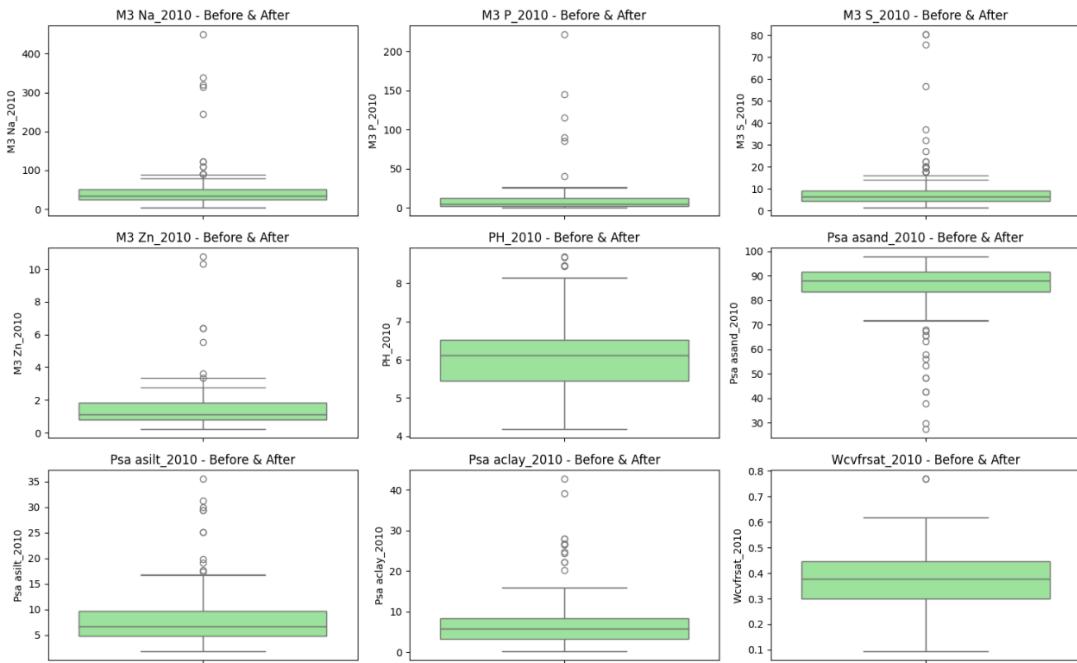
รูป 35 การกระจายตัวของข้อมูล ปี 2009 หลังการจัดการข้อมูลผิดปกติ (4)

Boxplots for 2010 - Part 1



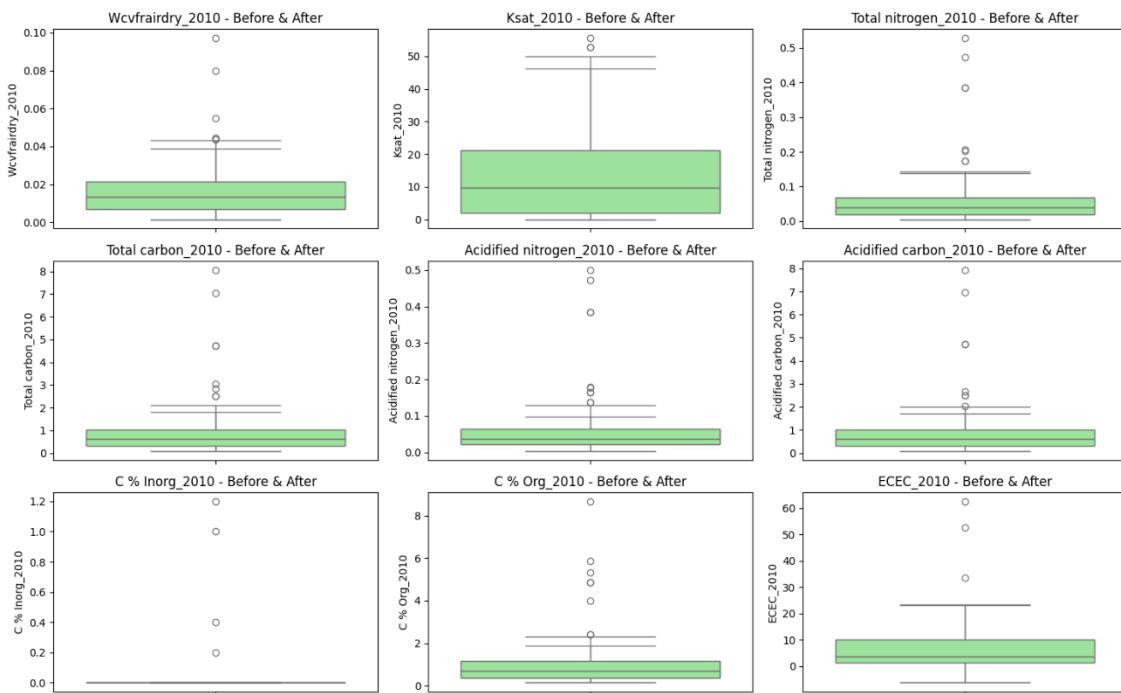
รูป 36 การกระจายตัวของข้อมูล ปี 2010 หลังการจัดการข้อมูลผิดปกติ (1)

Boxplots for 2010 - Part 2



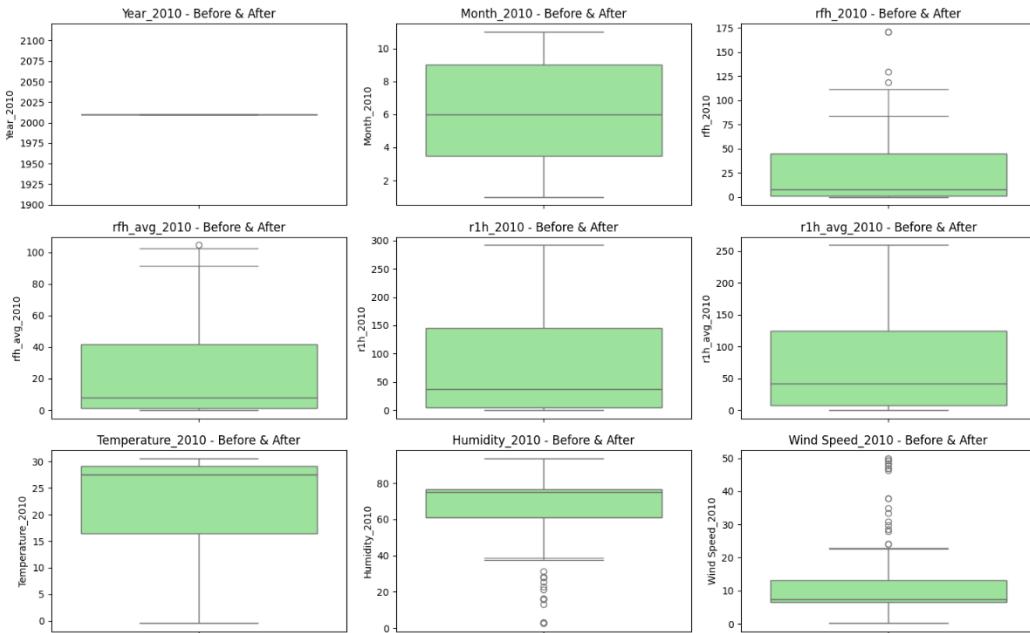
รูป 37 การกระจายตัวของข้อมูล ปี 2010 หลังการจัดการข้อมูลผิดปกติ (2)

Boxplots for 2010 - Part 3



รูป 38 การกระจายตัวของข้อมูล ปี 2010 หลังการจัดการข้อมูลผิดปกติ (3)

Boxplots for 2010 - Part 4



รูป 39 การกระจายตัวของข้อมูล ปี 2011 หลังการจัดการข้อมูลผิดปกติ (4)

เครื่องมือวัดทางวิทยาศาสตร์มีความถูกต้องแม่นยำและค่าผิดปกติเป็นค่าที่วัดได้จริงจากดิน ทางกลุ่มจะไม่มีการจัดการค่าผิดปกติและจะนำไปใช้เลย

3. Feature Engineering

หลังจากผ่านขั้นตอนการจัดการค่าผิดปกติ ถัดมาคือขั้นตอนการสร้างชุดข้อมูลใหม่ ผ่าน Feature Engineering โดยทางกลุ่มจะนำชุดข้อมูลปีต่างๆ มาคำนวณสำหรับการเพิ่มชุดข้อมูลเพื่อให้โมเดลเห็นแนวโน้มง่ายขึ้นเพื่อเพิ่มความแม่นยำของโมเดลโดยจะกำหนดชุดข้อมูลสำหรับคำนวณแสดงหาความสัมพันธ์และแนวโน้มของตัวแปร โดยมี 2 วิธีในการสร้างคุณลักษณะเชิงปฏิสัมพันธ์ ดังนี้

- เลือกนำชุดข้อมูลที่กำหนดมาหาอัตราส่วน (ratio) และผลคูณ (product) เพื่อนำไปสร้างชุดข้อมูลใหม่
- เลือกนำชุดข้อมูลที่กำหนดแบบเบเกเดี่ยวเพื่อนำมาหาค่าเฉลี่ย (mean) ของทั้ง 3 ปีและจึงนำมาหาค่าส่วนเบี่ยงเบนมาตรฐานเฉลี่ย (std) และนำมาระบบชุดข้อมูลใหม่ขึ้น

4. Feature Selection

การคัดเลือกว่าข้อมูลสำหรับหาแนวโน้มใดที่มีความเกี่ยวข้องและส่งผลต่อการคาดการณ์โดยตรงซึ่งทางกลุ่มจะให้วิธีหาข้อมูลดังกล่าวผ่าน Function SelectKBest(score_func=mutual_info_regression, k=7) ซึ่ง mutual_info_regression เป็นวิธีการคัดเลือกข้อมูลที่ความสัมพันธ์ระหว่างข้อมูลใช้คาดการณ์และข้อมูลทางแนวโน้มไม่มีความผันตรงหรือผกผันซึ่งกันและกันโดย Function จะเลือกชุดข้อมูล 7 ตัวแรกที่มีค่าความสัมพันธ์กับ Target สูงที่สุด โดยผลลัพธ์ที่ได้มีดังนี้

ตาราง 4 ชุดข้อมูลที่มีความสัมพันธ์กับ target

ID	7 ชุดข้อมูลที่มีความสัมพันธ์มากที่สุดกับ Target
EC	M3 Ca_Temperature_ratio_2009, M3 Na_Temperature_product_2009, M3 Na_Wind Speed_ratio_2009, PH_Temperature_product_2010, Temperature_M3 Na_product_2009, Temperature_PH_product_2010, Temperature_C % Org_ratio_2009
ECEC	M3 Al_PH_ratio_2010, M3 Al_Humidity_ratio_2010, M3 Mg_Total carbon_ratio_2011, M3 Na_M3 K_ratio_2011, M3 Na_Temperature_ratio_2009, Humidity_M3 Al_ratio_2010, M3 Al_std

ID	7 ชุดข้อมูลที่มีความสัมพันธ์มากที่สุดกับ Target
M3 B	M3 Ca_M3 S_product_2009, M3 Fe_M3 K_ratio_2009, M3 Mg_M3 Al_product_2009, M3 S_M3 Ca_product_2009, Ksat_Wind Speed_product_2009, Temperature_M3 B_ratio_2010, Wind Speed_Ksat_product_2009
M3 Ca	M3 Cu_Wcvfrairdry_ratio_2010, M3 Fe_Wcvfrsat_ratio_2010, M3 K_Temperature_product_2010, Psa asilt_M3 K_ratio_2009, Psa aclay_M3 K_ratio_2009, Temperature_M3 K_product_2010, Humidity_PH_ratio_2010
M3 Cu	M3 B_M3 Fe_ratio_2009, M3 B_PH_product_2009, M3 B_Temperature_ratio_2009, M3 Mn_M3 Zn_product_2010, M3 Zn_M3 Mn_product_2010, M3 Zn_Psa asand_product_2011, PH_M3 B_product_2009
M3 Fe	M3 AL_PH_ratio_2011, M3 B_M3 Mn_ratio_2009, M3 K_Total carbon_product_2011, M3 K_Acidified carbon_product_2011, Total carbon_M3 K_product_2011, Total carbon_Temperature_product_2011, Acidified carbon_M3 K_product_2011
M3 K	M3 AL_Wcvfrairdry_ratio_2011, M3 B_Acidified nitrogen_product_2009, M3 Na_M3 P_ratio_2011, M3 P_C % Org_ratio_2010, PH_M3 B_product_2009, Total nitrogen_M3 Mg_product_2011, Acidified nitrogen_M3 B_product_2009
M3 Mg	M3 B_Wcvfrsat_ratio_2011, M3 Cu_EC_ratio_2011, M3 K_Temperature_product_2010, Acidified carbon_Wcvfrairdry_ratio_2010, Temperature_M3 K_product_2010, Humidity_M3 K_ratio_2010, Wind Speed_mean
M3 Mn	r1h_avg_2011, M3 Fe_M3 Zn_ratio_2009, M3 Zn_C % Org_product_2011, Acidified carbon_M3 Zn_ratio_2011, C % Org_EC_ratio_2011, C % Org_M3 Zn_product_2011, C % Org_Total carbon_ratio_2009
M3 Na	M3 Ca_M3 Mg_ratio_2010, M3 Cu_M3 Zn_ratio_2011, M3 Mg_M3 Ca_ratio_2010, M3 Mn_M3 P_product_2009, M3 P_M3 Mn_ratio_2009, M3 P_M3 Mn_product_2009, M3 Zn_Psa asand_ratio_2010

ID	7 ชุดข้อมูลที่มีความสัมพันธ์มากที่สุดกับ Target
M3 S	M3 Fe_Psa asand_ratio_2009, Wcvfrsat_M3 Ca_ratio_2009, Total nitrogen_ECEC_product_2009, ECEC_Total nitrogen_product_2009, ECEC_Temperature_product_2009, Temperature_ECEC_product_2009, Humidity_M3 S_ratio_2009
M3 Zn	M3 Ca_M3 Al_ratio_2009, M3 Ca_Temperature_product_2009, M3 Zn_Humidity_product_2009, Psa asilt_EC_ratio_2009, Temperature_M3 Ca_product_2009, Humidity_M3 Zn_ratio_2009, Humidity_M3 Zn_product_2009
pH	M3 Al_PH_ratio_2009, M3 Al_C % Org_product_2009, M3 Al_Humidity_ratio_2010, M3 Na_M3 Mn_ratio_2009, PH_M3 Fe_ratio_2010, C % Org_M3 Al_product_2009, Temperature_M3 Al_ratio_2010
Psa asand	M3 B_Psa asand_product_2009, M3 K_M3 Fe_ratio_2010, M3 S_M3 B_ratio_2010, Psa asand_M3 B_product_2009, Psa asand_Temperature_product_2011, Temperature_Psa asand_product_2011, Wcvfrsat_std
Psa asilt	M3 B_M3 S_product_2011, M3 Mg_Psa asand_product_2010, M3 Na_Total nitrogen_product_2011, M3 S_M3 B_product_2011, Ksat_Total carbon_ratio_2011, Total nitrogen_M3 Na_product_2011, Total carbon_M3 K_ratio_2011
Psa aclay	M3 B_2009, M3 B_PH_product_2009, M3 B_C % Inorg_ratio_2009, M3 B_C % Inorg_ratio_2011, M3 Na_Psa asand_ratio_2011, M3 Na_C % Inorg_ratio_2011, PH_M3 B_product_2009
Wcvfrsat	M3 B_PH_product_2009, M3 Cu_M3 Na_product_2010, M3 Na_M3 Cu_product_2010, M3 S_Wcvfrairdry_ratio_2009, PH_M3 B_product_2009, Psa asilt_M3 P_ratio_2009, M3 Al_std
Ksat	M3 Ca_EC_ratio_2010, PH_Psa asand_product_2009, Psa asilt_PH_ratio_2010, Psa asilt_ECEC_product_2010, Psa aclay_EC_ratio_2009, ECEC_Psa asilt_product_2010, rfh_avg_std

ID	7 ชุดข้อมูลที่มีความสัมพันธ์มากที่สุดกับ Target
Total carbon	M3 Fe_M3 K_ratio_2011, PH_Wcvfrsat_ratio_2009, Psa aclay_Total carbon_product_2009, Wcvfrsat_C % Org_product_2009, Total carbon_Psa aclay_product_2009, Acidified carbon_M3 Mg_ratio_2009, C % Org_Wcvfrsat_product_2009
C % Inorg	M3 P_C % Inorg_product_2009, Acidified nitrogen_C % Inorg_product_2010, C % Inorg_M3 S_product_2010, C % Inorg_M3 Zn_product_2010, C % Inorg_Acidified nitrogen_ratio_2010, C % Inorg_ECEC_product_2010, C % Inorg_Humidity_product_2010
C % Org	M3 Al_M3 Mg_ratio_2009, PH_Humidity_ratio_2009, Psa asand_Total carbon_ratio_2010, Psa asand_C % Org_ratio_2010, Wcvfrsat_Wind Speed_product_2009, C % Org_M3 Na_ratio_2009, Humidity_PH_ratio_2009

5. Standardization

ทางกลุ่มจะปรับสเกลของชุดข้อมูลทุกตัวให้อยู่ในช่วงที่เท่ากันโดยโมเดลที่จะต้องผ่านขั้นตอนนี้จะมีแค่ SVR และ Neural Network เท่านั้น

```
scaler = StandardScaler()  
X_scaled = scaler.fit_transform(df_combined_imputed)
```

รูป 40 โปรแกรมการใช้ StandardScaler(Standardization)

3.2.3 Model Construction

1. แบ่งชุดข้อมูล Training Set and Testing Set

โดยการกำหนดให้ปี 2009 - 2011 เป็น Training Set ซึ่งจะไม่นำข้อมูลที่ซื้อเหมือนกันมารวมกันแต่จะแยกข้อมูลกันตามรายปี แล้วนำมาเรียงกันและในส่วนของ Testing Set จะใช้ข้อมูลที่ต้องการจะคาดการณ์ของปี 2012 มาเป็นตัวเปรียบเทียบความแม่นยำกับค่าที่คาดการณ์ได้



รูป 41 ลักษณะการแบ่งข้อมูลในการเทรนโมเดล

2. การเลือกโมเดล Machine Learning

ทางกลุ่มเลือกทดลองใช้โมเดล 6 ประเภท ได้แก่ Random Forest, Support Vector Regression (SVM), Decision Tree, Gradient Boosting, Extreme Gradient Boosting และ Neural Networks

โดยโมเดลทุกแบบจะถูกนำมาทดลองทั้งกรณีที่จัดการกับค่าผิดปกติและไม่จัดการเพื่อหาว่าค่าผิดปกติเป็นค่าที่เครื่องมือวัดทางวิทยาศาสตร์ผิดพลาดหรือไม่โดยจะทำพร้อมกับการเทรนโมเดลเพื่อหาความแม่นยำของแต่ละโมเดลร่วมไปด้วย โดยทางกลุ่มจะใช้วิธีการหาค่าต่างๆ ของตัวแปรในโมเดลที่เหมาะสมและสามารถเพิ่มความ

แม่นยำต่อการคาดการณ์มากที่สุด (Hyperparameter) โดยวิธีที่ใช้คือ Grid Search หรือการนำค่าที่กลุ่มกำหนดไว้ในแต่ละตัวแปรต่างๆ มาทดลองใช้กับโมเดลแล้วจะเลือกค่าที่ให้ความแม่นยำสูงที่สุด

3. การประเมินผลโมเดล

ผลลัพธ์ที่ได้จากการเทรนโมเดลจะมี 4 อย่างดังนี้

1. R2 score หรือค่าความสามารถในการเข้าใจค่าที่ใช้เทรนโมเดล ยิ่งค่าเข้าใกล้ 1 โมเดลก็ยิ่งสามารถเข้าใจค่าต่างๆ ในชุดข้อมูลได้มาก
2. MAE หรือ ระยะห่าง(Absolute Error) ระหว่างค่าจริงและค่าที่คาดการณ์
3. RMSE หรือค่าต่าง(Error) ระหว่างค่าจริงและค่าคาดการณ์
4. ผลลัพธ์ของสารเคมีในดินต่างๆ(Target) ที่ผลลัพธ์คาดการณ์ออกมา

ตาราง 5 การเปรียบเทียบโมเดล

Model Output	Meaning
R2 score	ค่าความสามารถในการเข้าใจค่าที่ใช้เทรนโมเดล ยิ่งค่าเข้าใกล้ 1 โมเดลก็ยิ่งสามารถเข้าใจค่าต่างๆ ในชุดข้อมูลได้มาก
MAE (Mean Absolute Error)	ระยะห่าง(Absolute Error) ระหว่างค่าจริงและค่าที่คาดการณ์
RMSE (Root Mean Square Error)	ค่าต่าง(Error) ระหว่างค่าจริงและค่าคาดการณ์
Predicted Values	ผลลัพธ์ของสารเคมีในดินต่างๆ(Target) ที่ผลลัพธ์คาดการณ์ออกมา

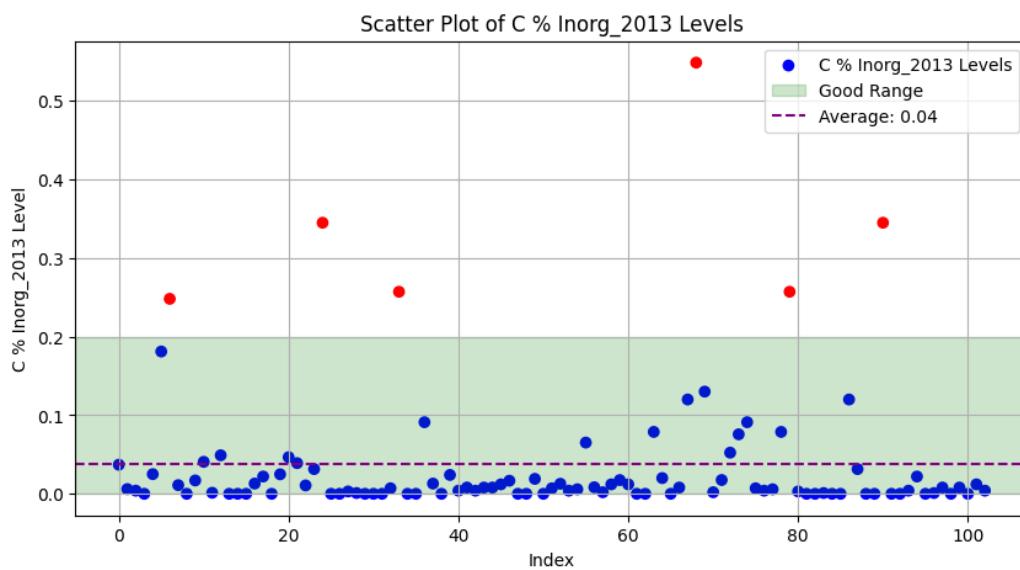
โดย R2, MAE และ RMSE จะถูกคำนวณไว้คร่าวๆ เพื่อประเมินว่าโมเดลได้ทึมีการจัดการค่าผิดปกติ/ไม่จัดการค่าผิดปกติ ให้ความแม่นยำมากที่สุด

4. หลังจากประเมินความแม่นยำของโมเดลทั้งหมดผ่าน R², MAE และ RMSE เรียบร้อยทางกลุ่มก็จะเลือกโมเดลที่มี 3 ค่าดังกล่าวที่ดีที่สุดมาคาดการณ์ผลลัพธ์ เมื่อคาดการณ์เรียบร้อยแล้ว ผลลัพธ์ที่ได้จะถูกนำมาประเมินผ่านดัชนีคุณภาพสารเคมีในดินและหาค่าเฉลี่ยรายปีของสารเคมีในดินแต่ละอย่างแล้วแสดงการประเมินให้แก่ผู้ใช้งานรูปภาพต่อไปนี้

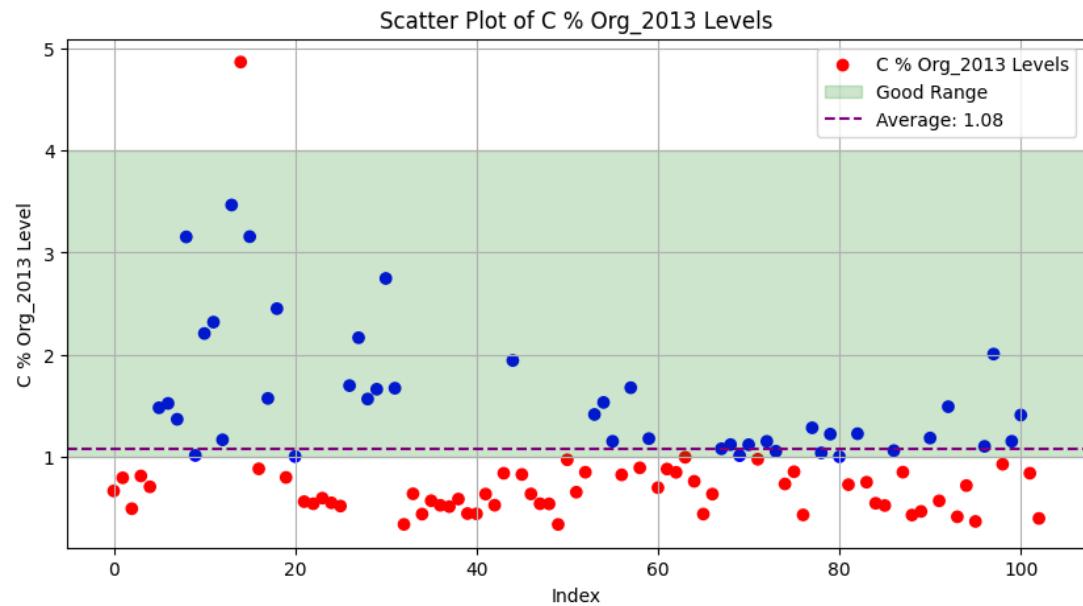
Soil Feature Averages and Classifications		
Feature	Average Value	Classification
C % Inorg_2012 (%)	0.03	Normal C % Inorg_2012 Value in Soil
C % Org_2012 (%)	1.00	Normal C % Org_2012 Deficient Soil
ECEC_2012 (cmol/kg)	7.45	Normal ECEC_2012 Value in Soil
EC_2012 (dS/m)	0.14	Normal EC_2012 Value in Soil
Ksat_2012 (cm/hr)	11.77	Normal Drainage Soil
M3 Al_2012 (mg/kg)	805.46	M3 Al_2012 Toxic Soil
M3 B_2012 (mg/kg)	0.55	Normal M3 B_2012 Value in Soil
M3 Ca_2012 (mg/kg)	1181.33	Normal M3 Ca_2012 Value in Soil
M3 Cu_2012 (mg/kg)	1.09	Normal M3 Cu_2012 Value in Soil
M3 Fe_2012 (mg/kg)	104.77	Normal M3 Fe_2012 Value in Soil
M3 K_2012 (mg/kg)	172.55	Normal M3 K_2012 Value in Soil
M3 Mg_2012 (mg/kg)	231.55	Normal M3 Mg_2012 Value in Soil
M3 Mn_2012 (mg/kg)	84.08	Normal M3 Mn_2012 Value in Soil
M3 Na_2012 (mg/kg)	110.59	M3 Na_2012 Toxic Soil
M3 P_2012 (mg/kg)	9.36	M3 P_2012 Deficient Soil
M3 S_2012 (mg/kg)	21.16	M3 S_2012 Toxic Soil
M3 Zn_2012 (mg/kg)	1.50	Normal M3 Zn_2012 Value in Soil
pH_2012 (0)	5.98	Neutral Soil
Psa_aclay_2012 (%)	7.33	Poor Drainage Soil
Psa_asand_2012 (%)	84.70	Good Fertility and Drainage in Soil
Psa_asilt_2012 (%)	7.94	Low Fertility in Soil
Total carbon_2012 (%)	0.98	Total carbon_2012 Deficient Soil
Total nitrogen_2012 (%)	0.07	Normal Total nitrogen_2012 Value in Soil
WcVfrat_2012 (m3/m3)	0.36	Normal Water Saturation in Soil

รูป 42 แสดงหน้าต่างระบบชื่อ Target, หน่วย ค่าเฉลี่ยรายปี และผลการประเมิน

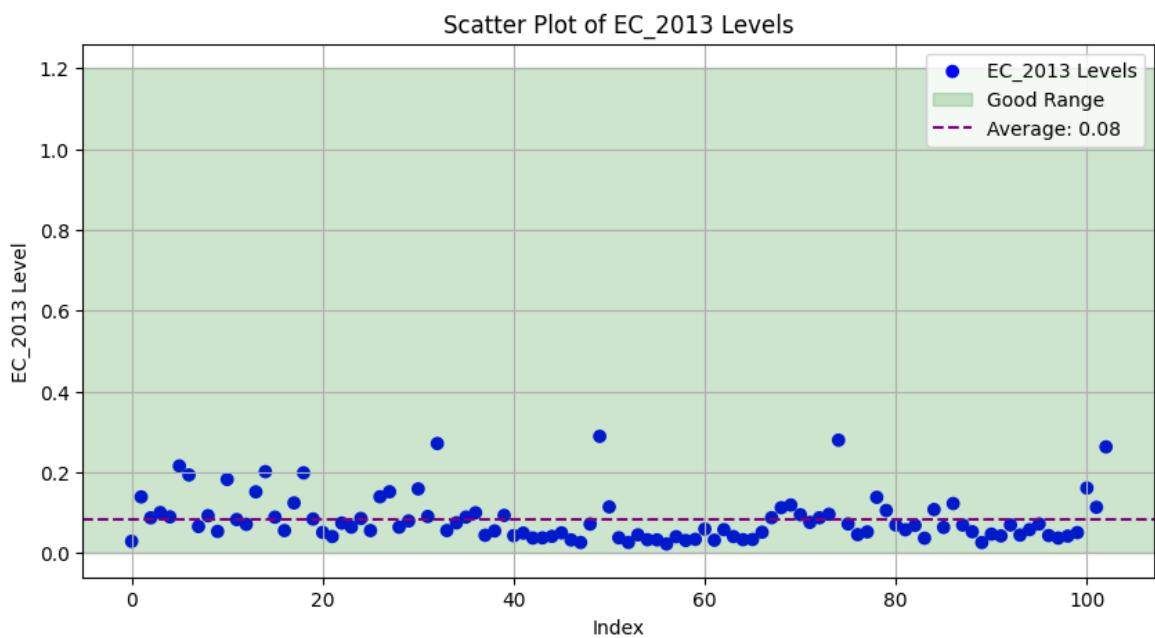
นอกเหนือจากการแสดงผลกราฟประเมินแล้วโปรแกรมยังแสดงค่าของข้อมูลในรูปแบบ Scatter Plot เพื่อนำมาวิเคราะห์เชิงลึกต่อไปถึงการกระจายตัวของข้อมูลที่คาดการณ์



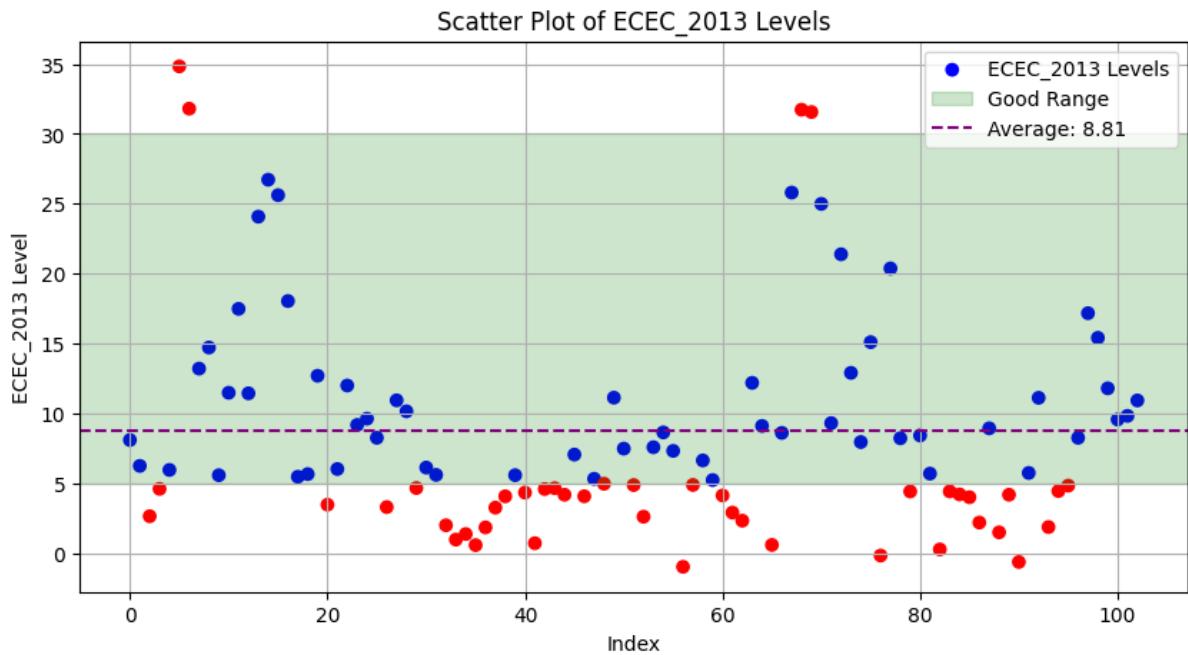
รูป 43 แสดงค่าของข้อมูล C % Inorg ที่มีการระบุช่วงที่เป็นผลดีต่อดิน



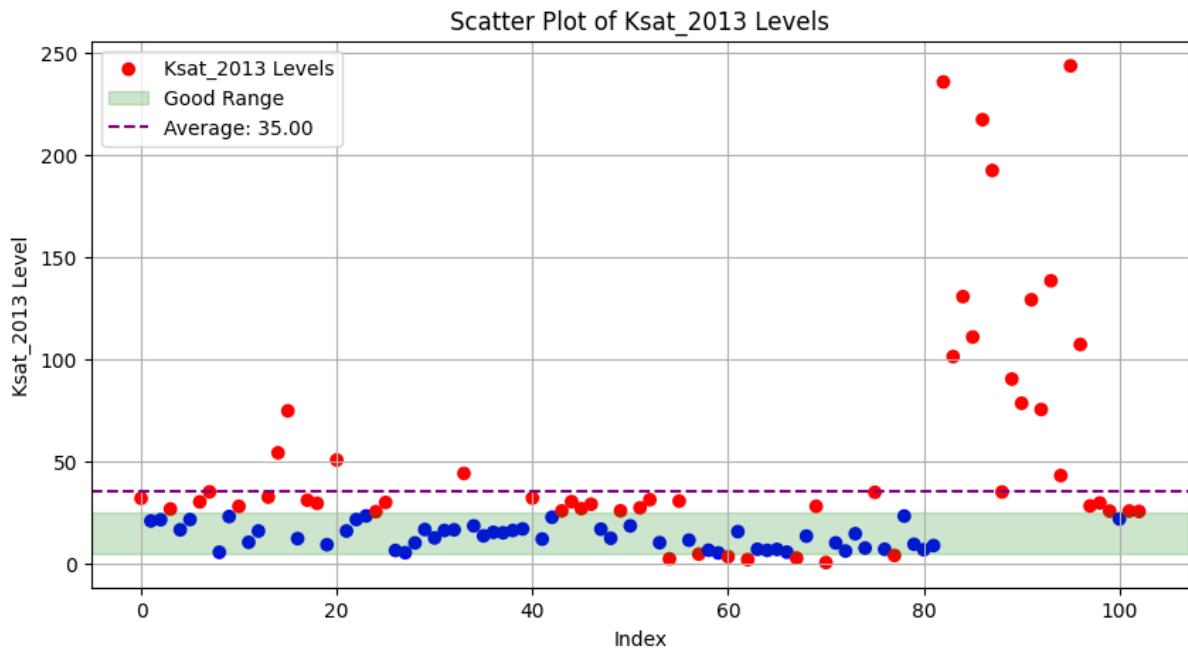
รูป 44 แสดงค่าของข้อมูล C % Org ที่มีการระบุช่วงที่เป็นผลดีต่อ din



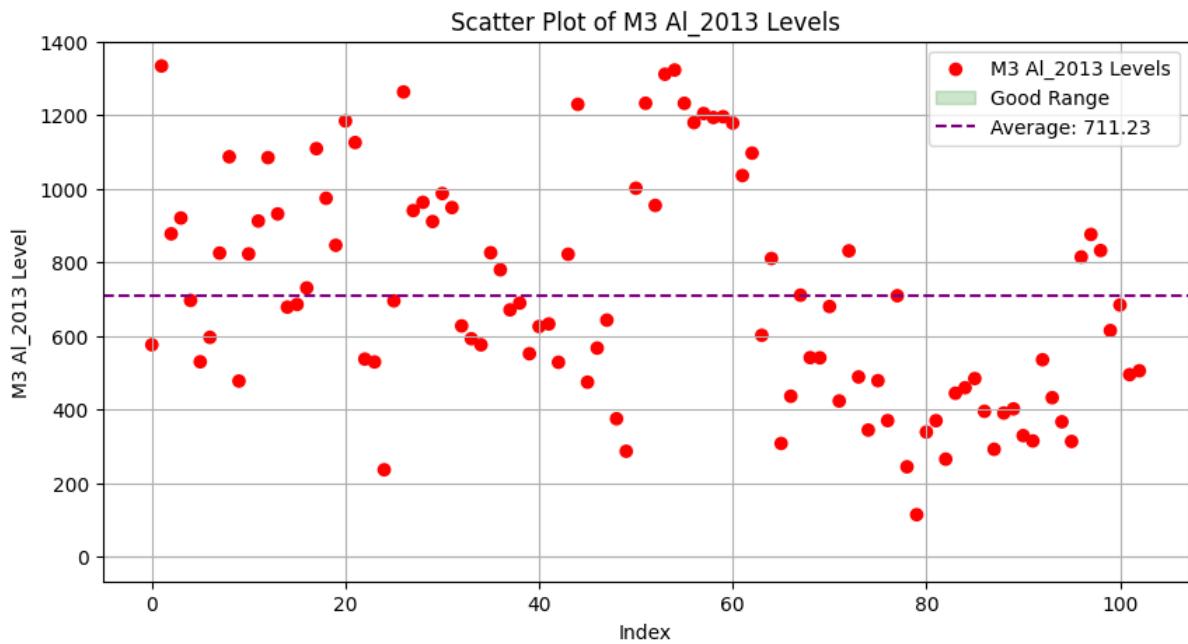
รูป 45 แสดงค่าของข้อมูล EC ที่มีการระบุช่วงที่เป็นผลดีต่อ din



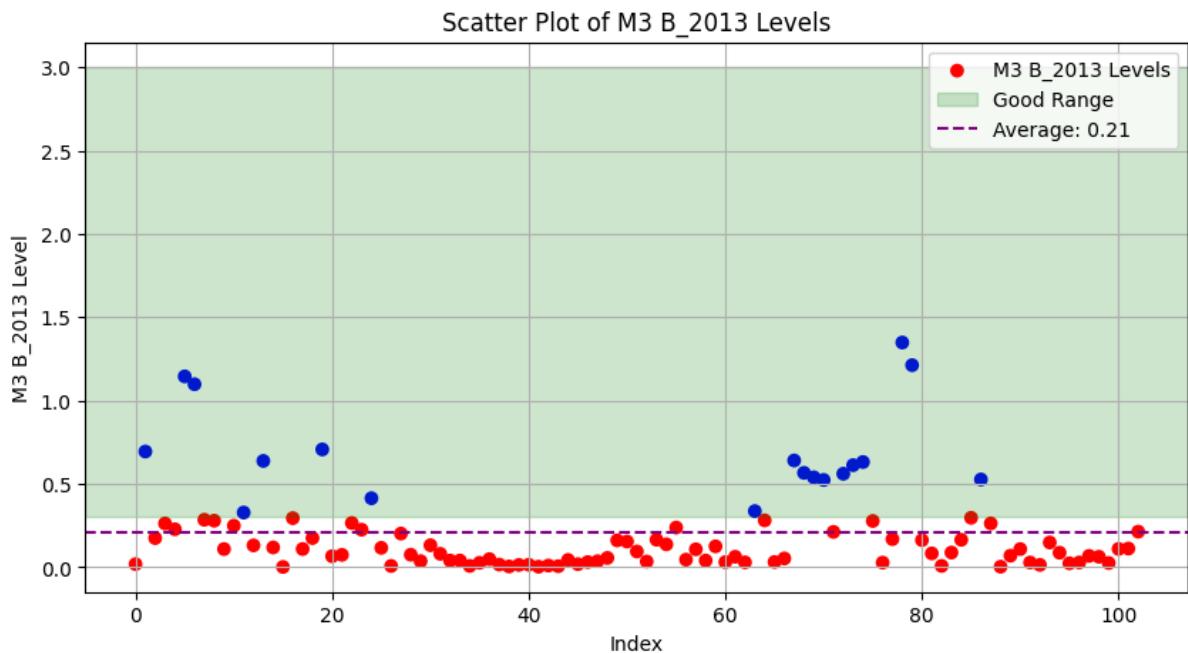
รูป 46 แสดงค่าของข้อมูล ECEC ที่มีการระบุช่วงที่เป็นผลดีต่อ din



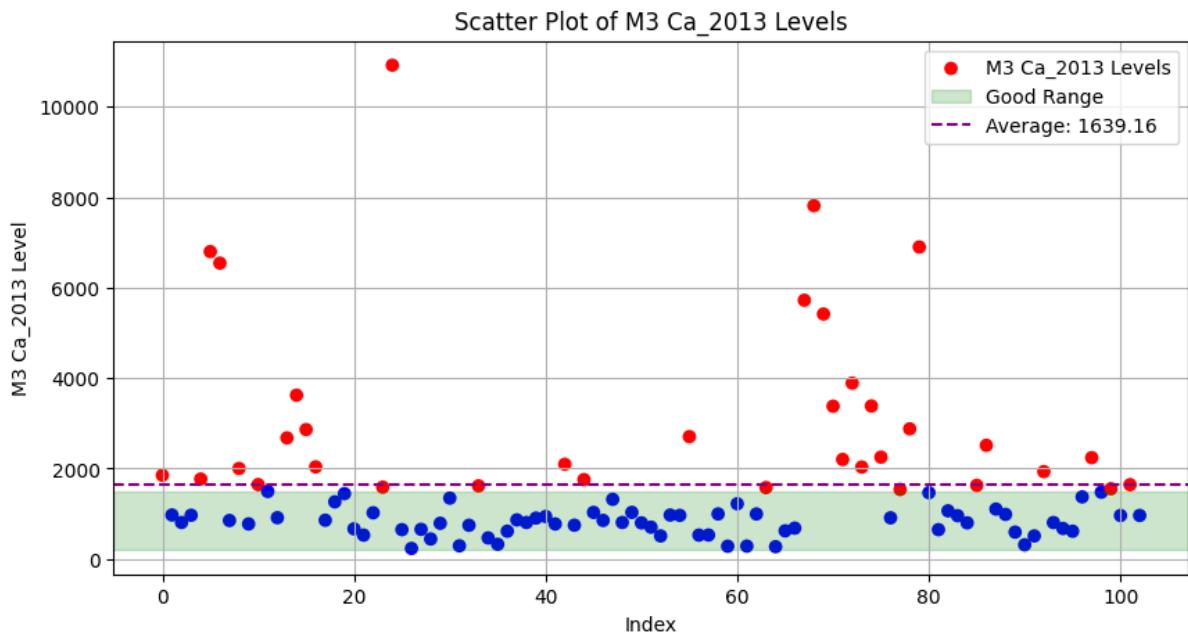
รูป 47 แสดงค่าของข้อมูล Ksat ที่มีการระบุช่วงที่เป็นผลดีต่อ din



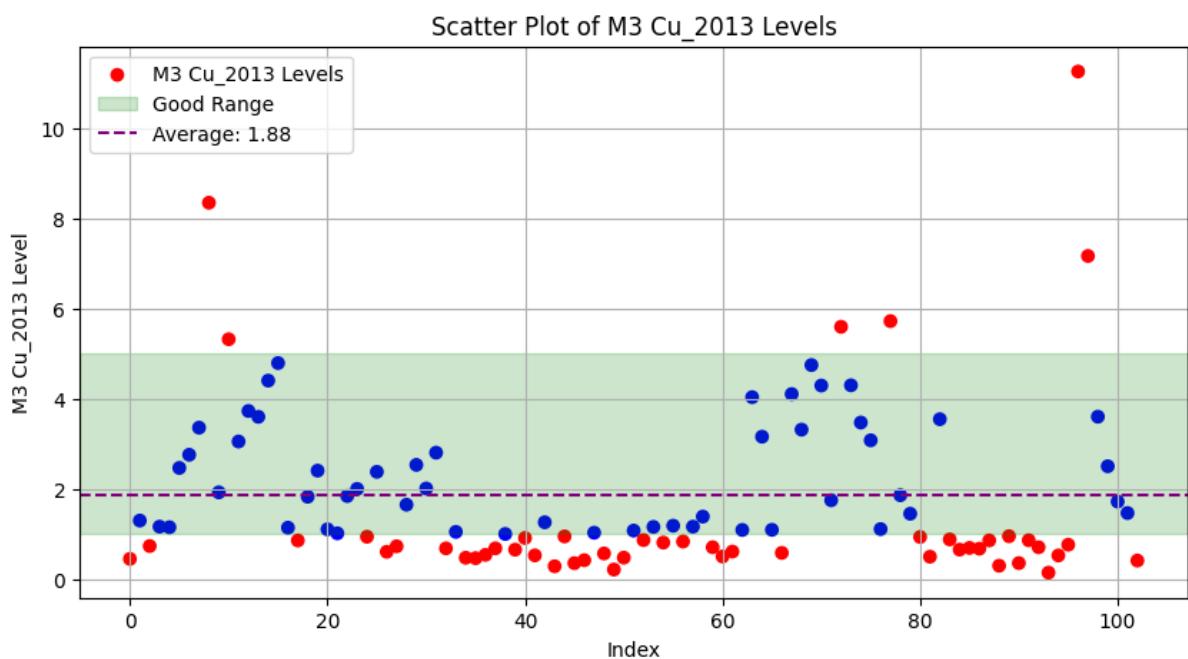
รูป 48 แสดงค่าของข้อมูล M3 AI ที่มีการระบุช่วงที่เป็นผลดีต่อдин



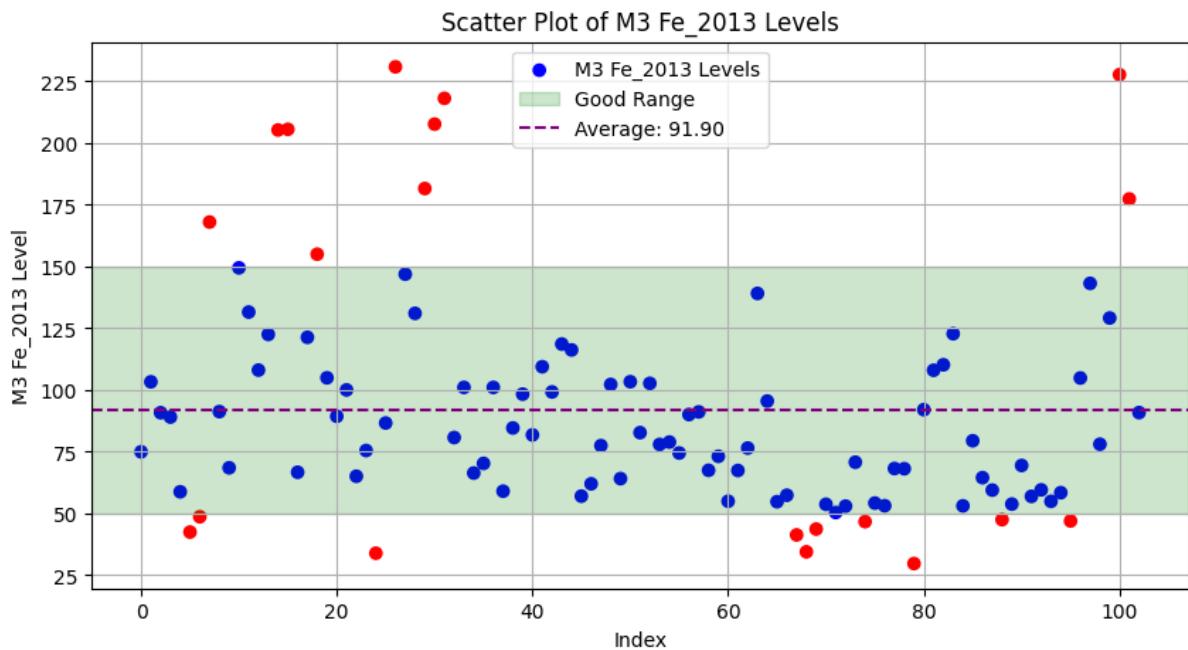
รูป 49 แสดงค่าของข้อมูล M3 B ที่มีการระบุช่วงที่เป็นผลดีต่อдин



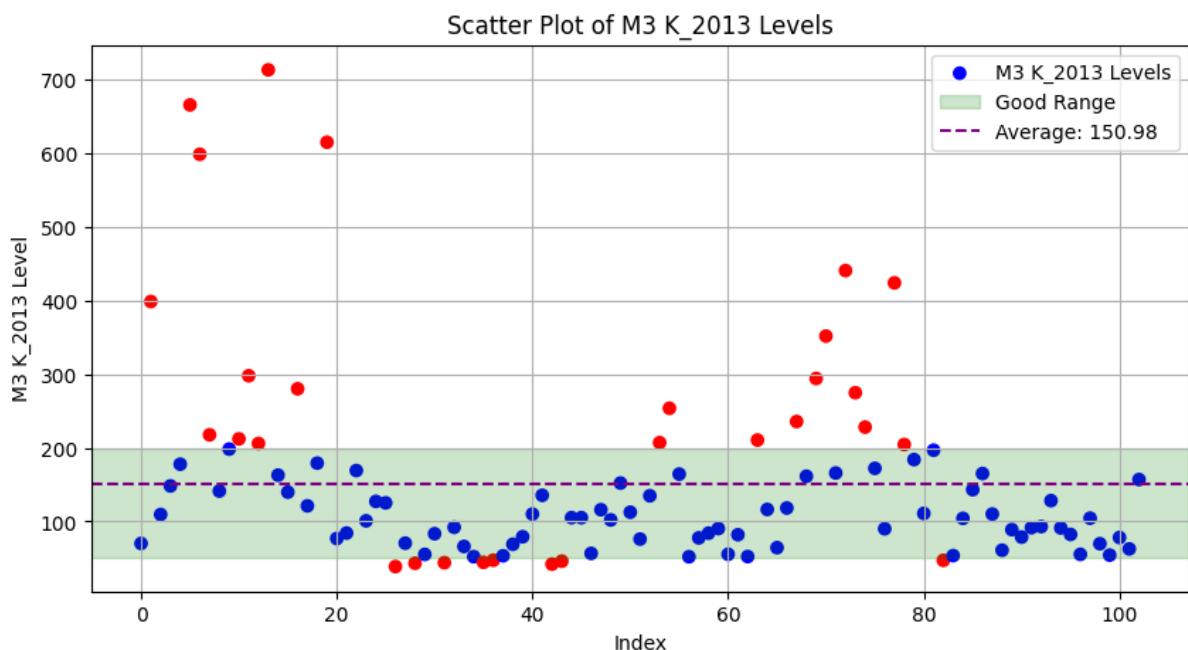
รูป 50 แสดงค่าของข้อมูล M3 Ca ที่มีการระบุช่วงที่เป็นผลดีต่อ din



รูป 51 แสดงค่าของข้อมูล M3 Cu ที่มีการระบุช่วงที่เป็นผลดีต่อ din

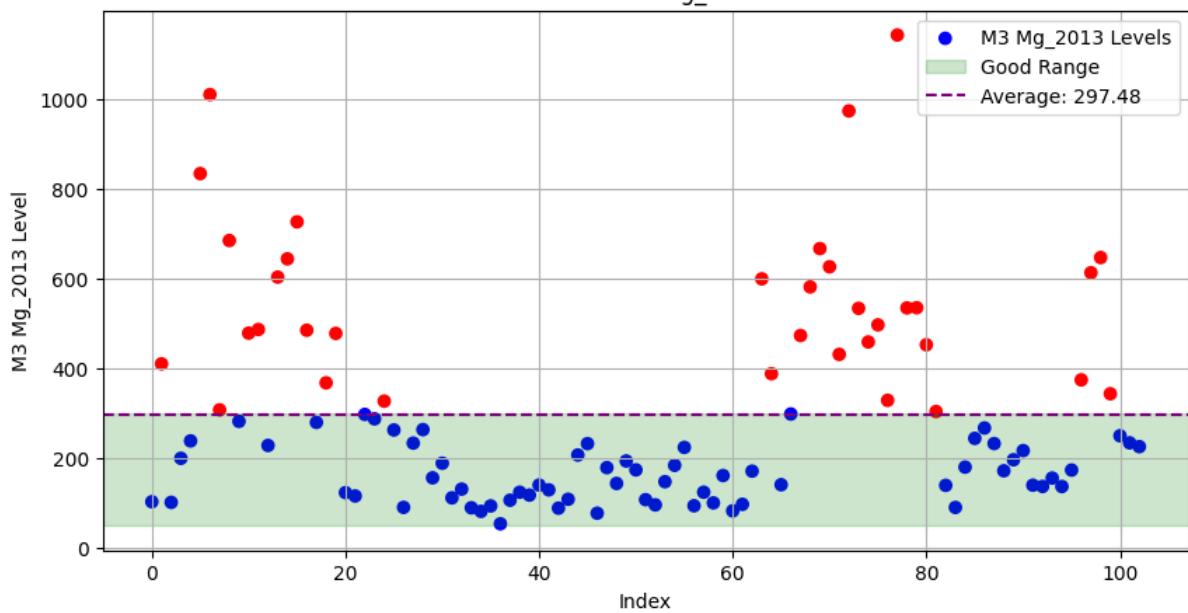


รูป 52 แสดงค่าของข้อมูล M3 Fe ที่มีการระบุช่วงที่เป็นผลดีต่อ din



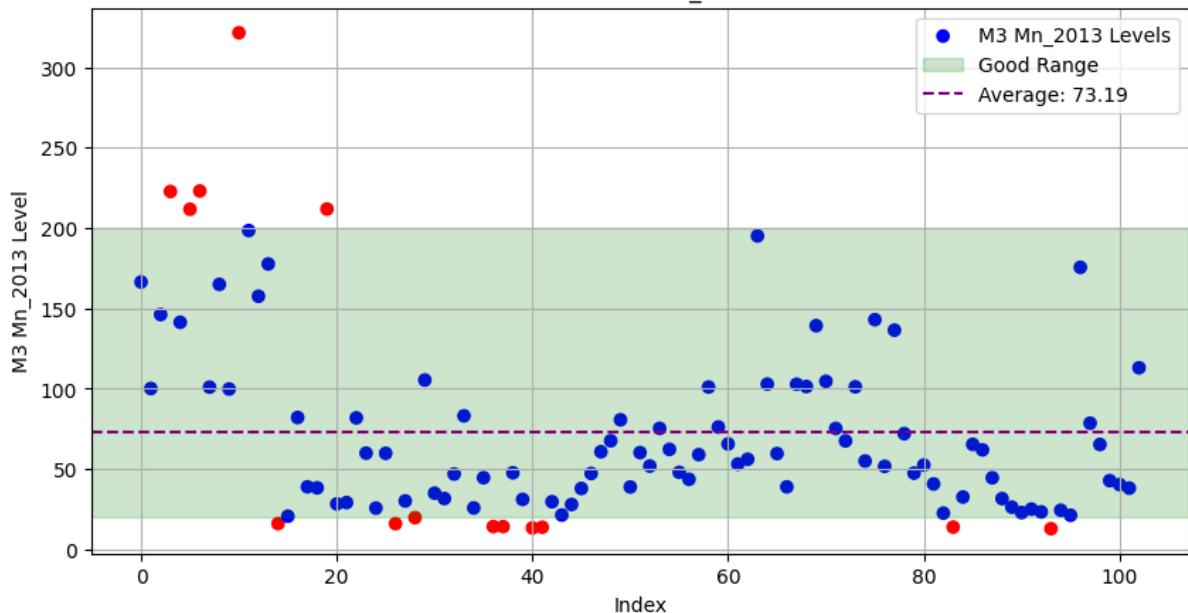
รูป 53 แสดงค่าของข้อมูล M3 K ที่มีการระบุช่วงที่เป็นผลดีต่อ din

Scatter Plot of M3 Mg_2013 Levels

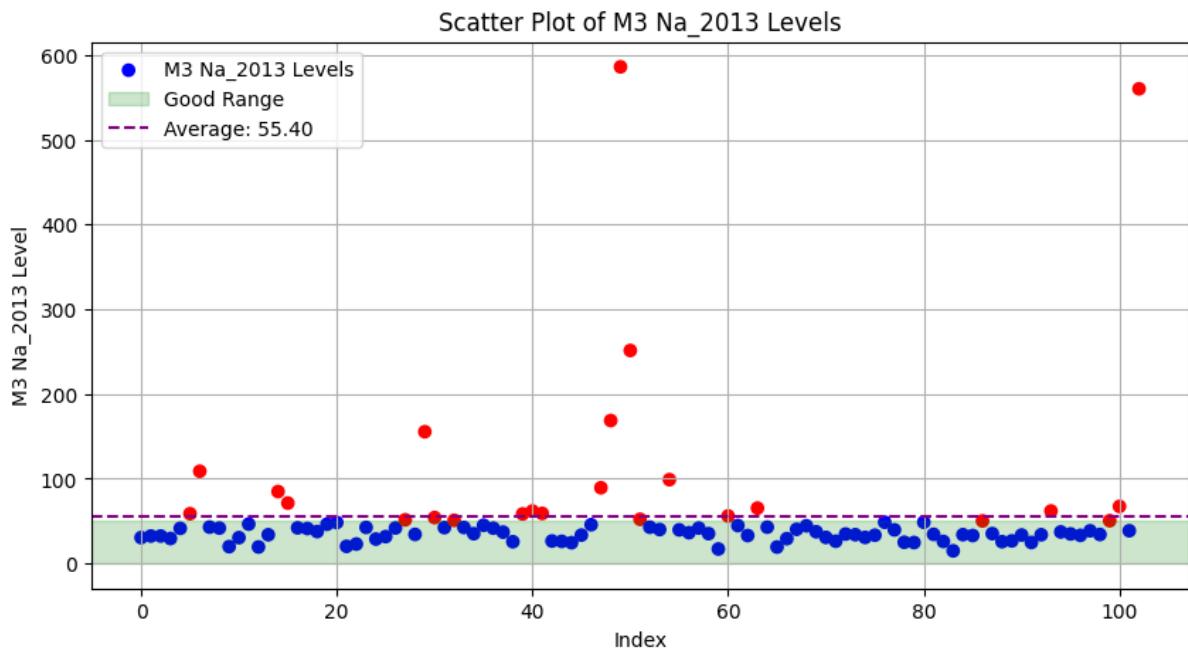


รูป 54 แสดงค่าของข้อมูล M3 Mg ที่มีการระบุช่วงที่เป็นผลดีต่อдин

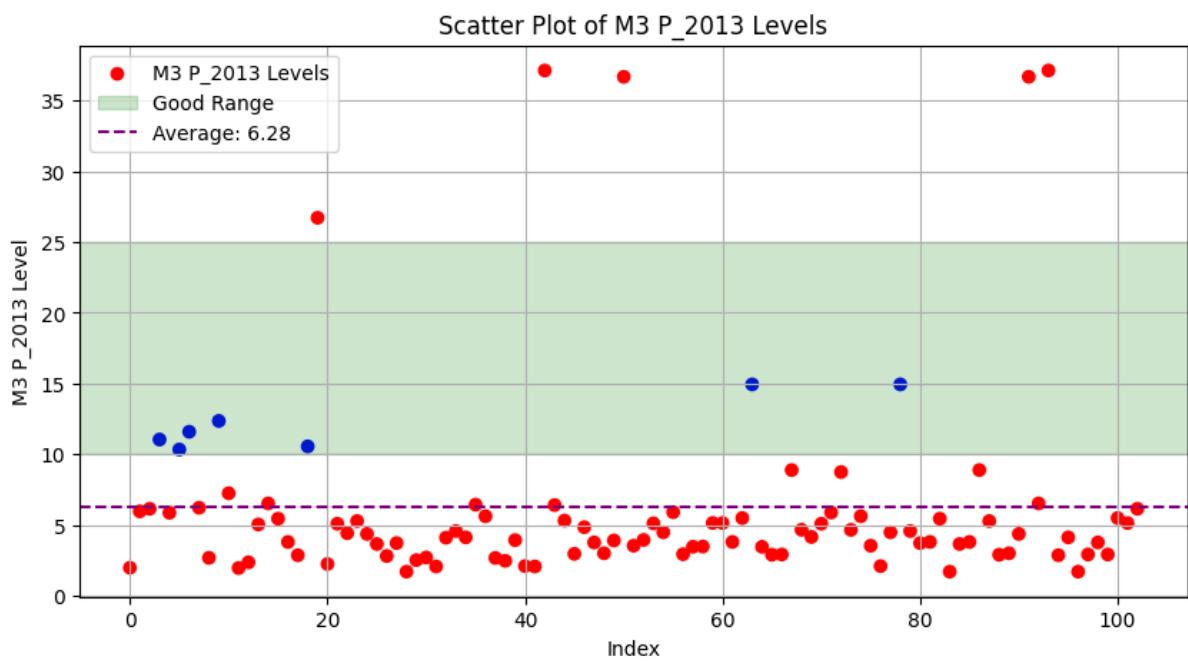
Scatter Plot of M3 Mn_2013 Levels



รูป 55 แสดงค่าของข้อมูล M3 Mn ที่มีการระบุช่วงที่เป็นผลดีต่อдин

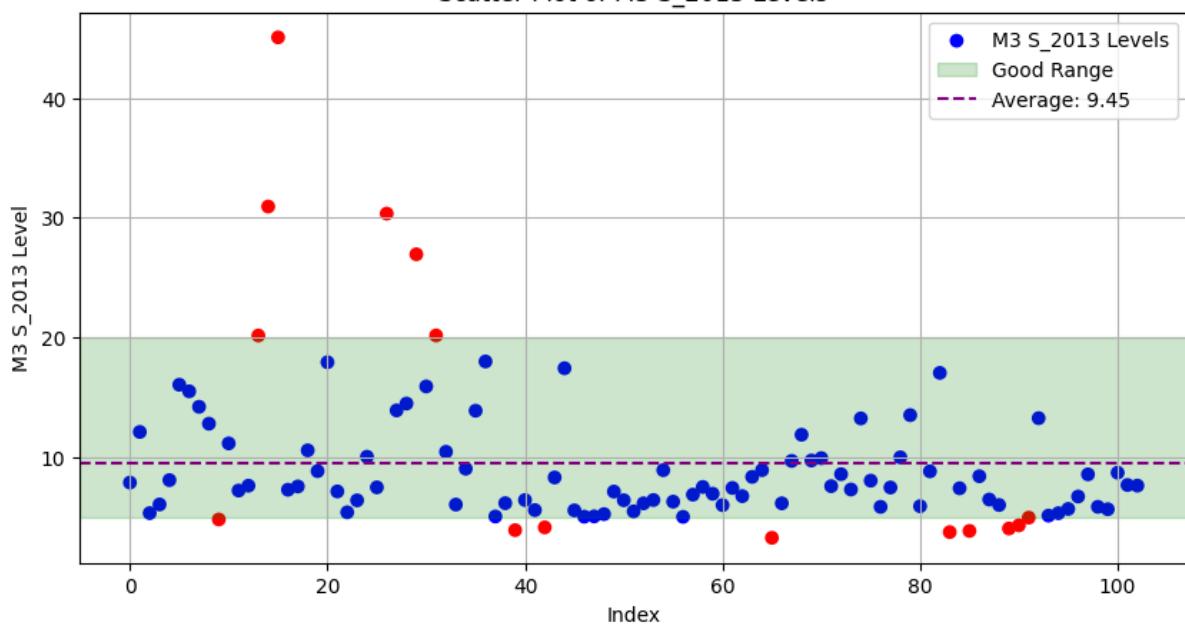


รูป 56 แสดงค่าของข้อมูล M3 Na ที่มีการระบุช่วงที่เป็นผลดีต่อдин



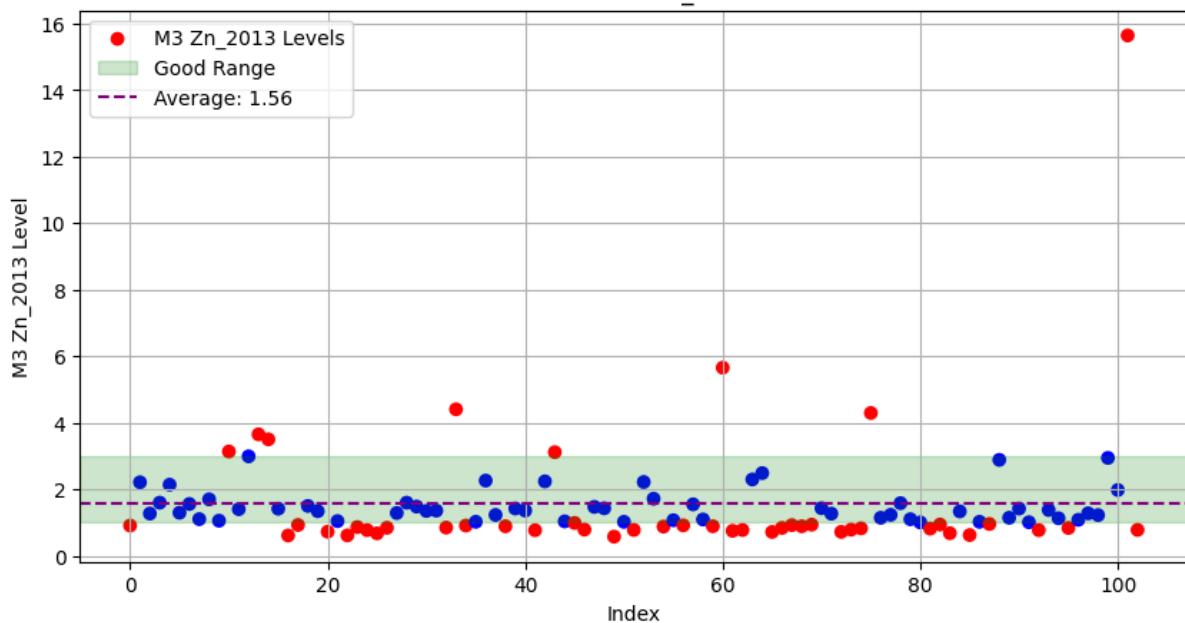
รูป 57 แสดงค่าของข้อมูล M3 P ที่มีการระบุช่วงที่เป็นผลดีต่อдин

Scatter Plot of M3 S_2013 Levels

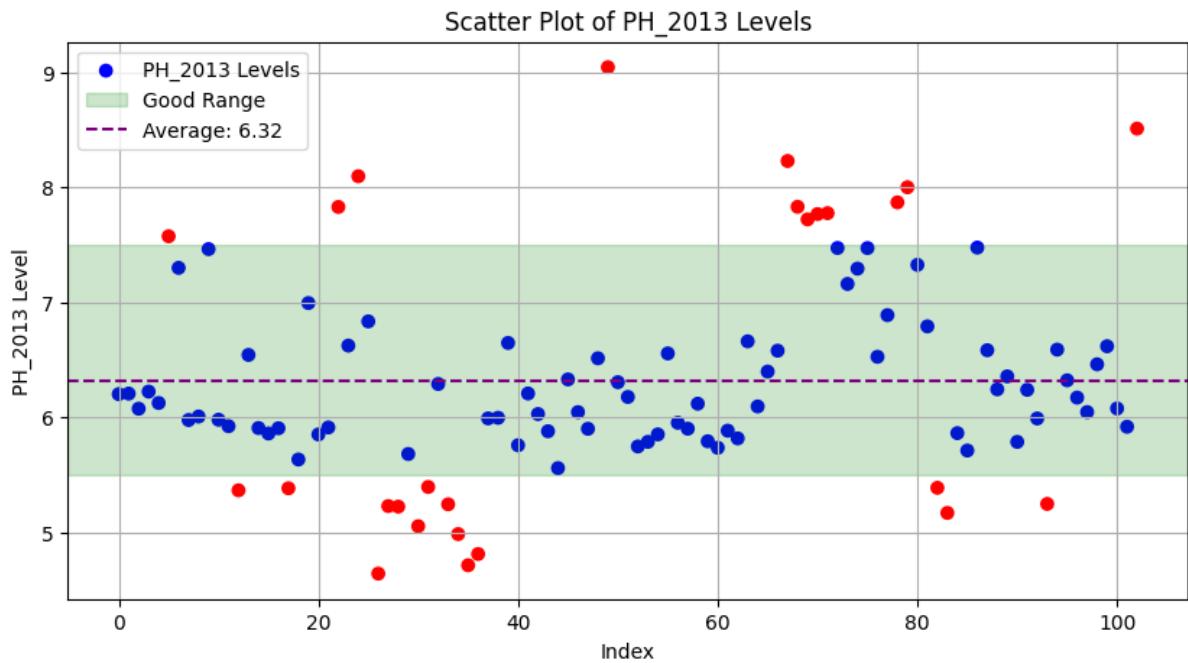


รูป 58 แสดงค่าของข้อมูล M3 S ที่มีการระบุช่วงที่เป็นผลดีต่อ din

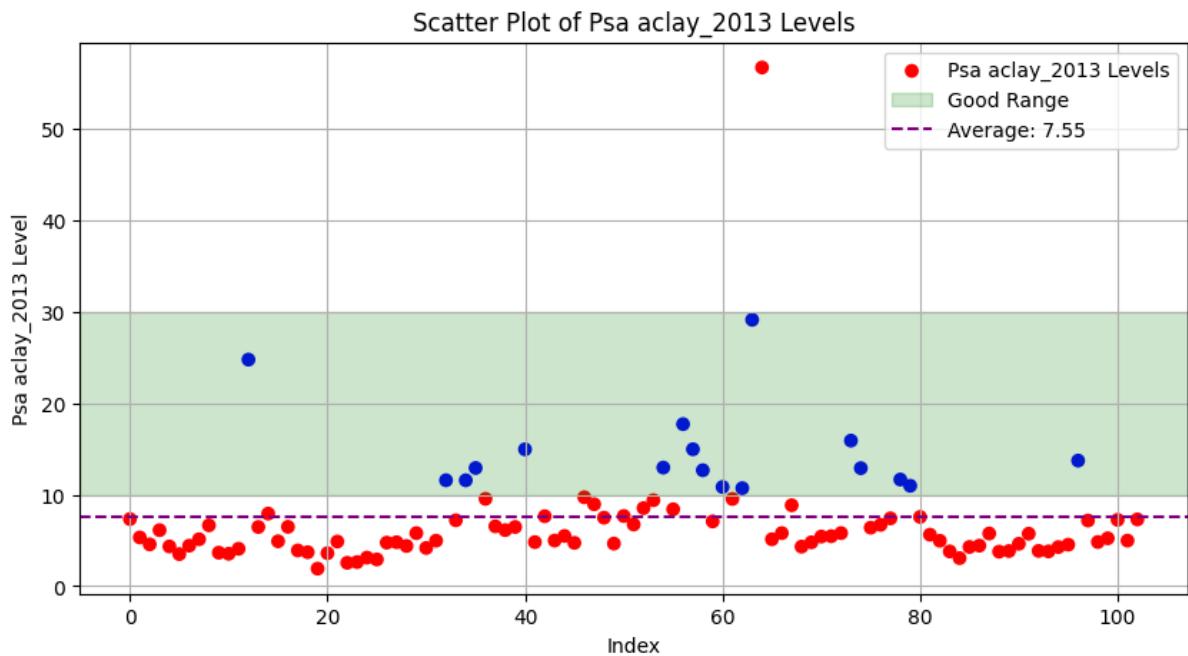
Scatter Plot of M3 Zn_2013 Levels



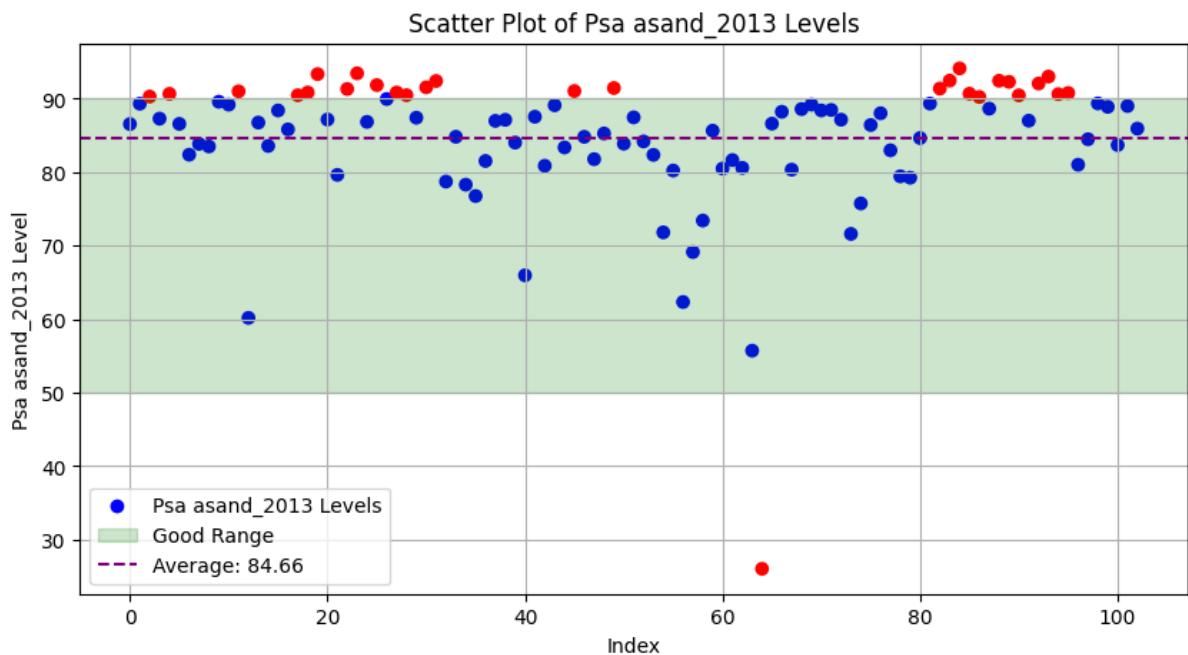
รูป 59 แสดงค่าของข้อมูล M3 Zn ที่มีการระบุช่วงที่เป็นผลดีต่อ din



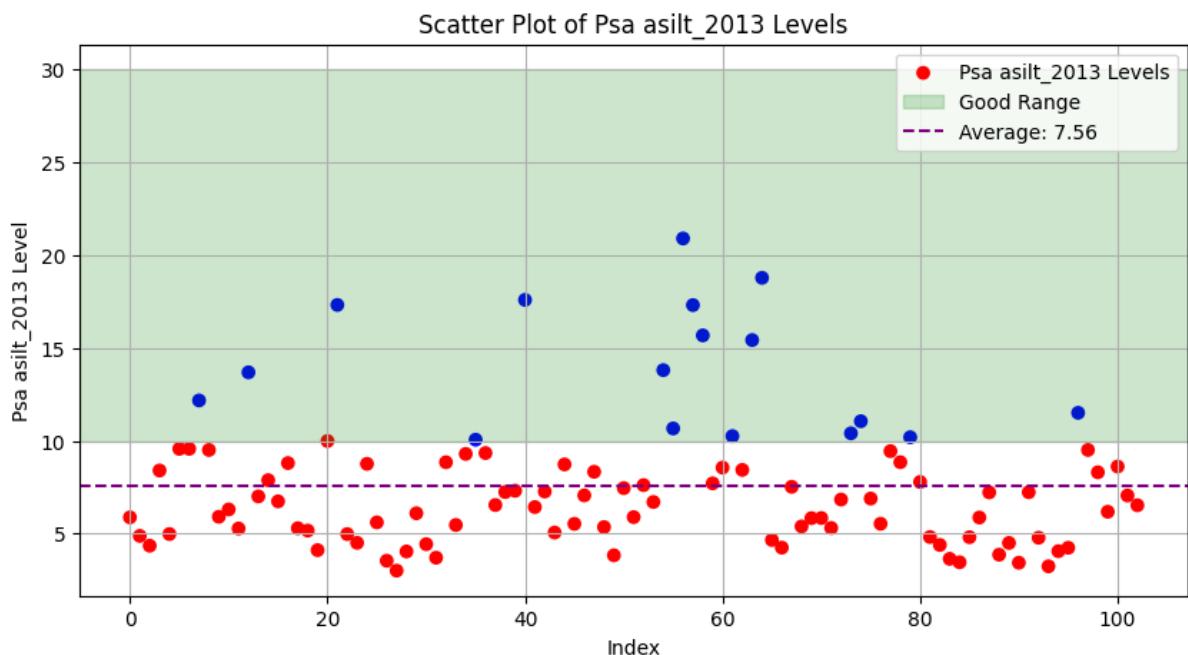
รูป 60 แสดงค่าของข้อมูล PH ที่มีการระบุช่วงที่เป็นผลดีต่อ din



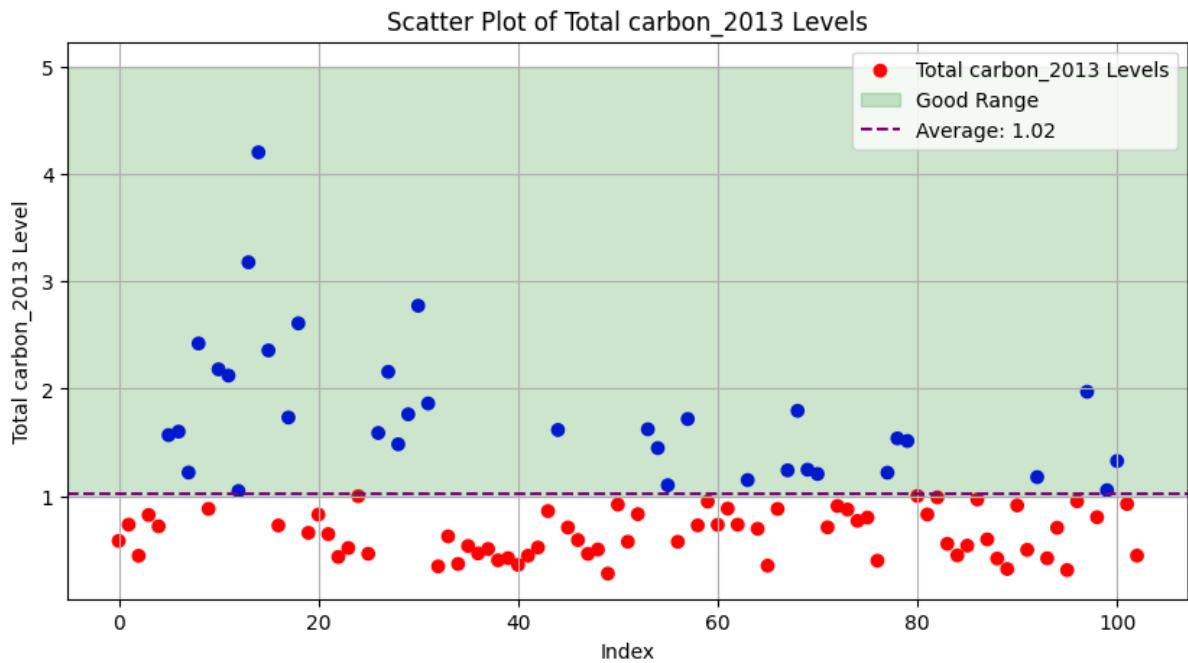
รูป 61 แสดงค่าของข้อมูล Psa aclay ที่มีการระบุช่วงที่เป็นผลดีต่อ din



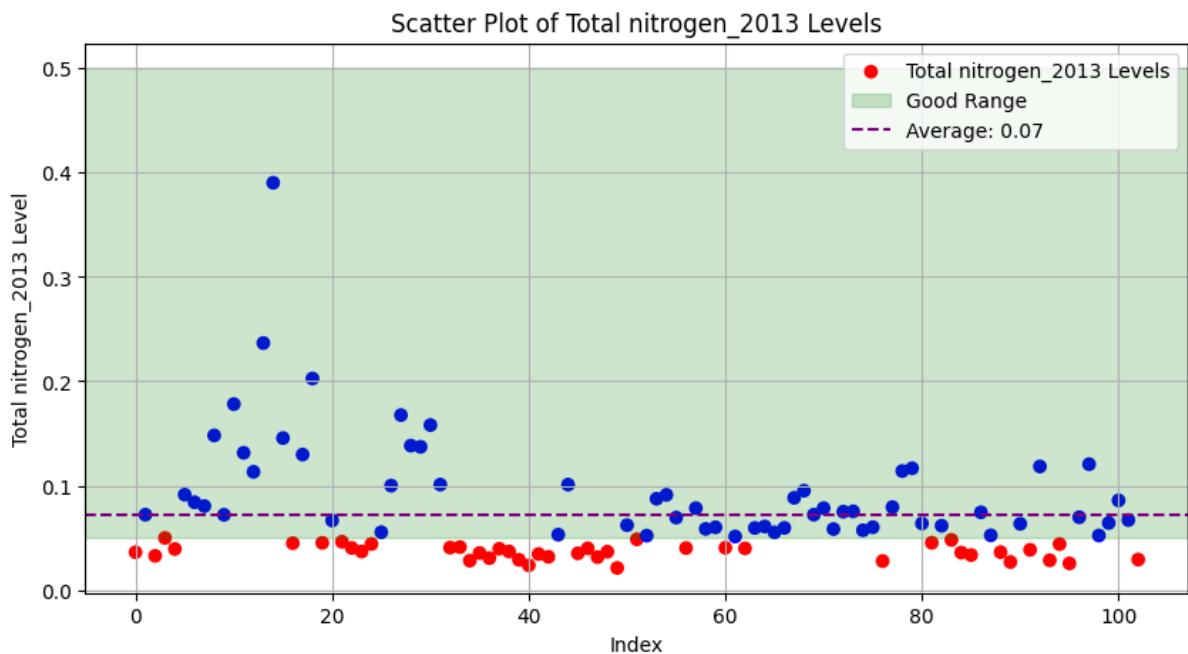
รูป 62 แสดงค่าของข้อมูล Psa asand ที่มีการระบุช่วงที่เป็นผลดีต่อ din



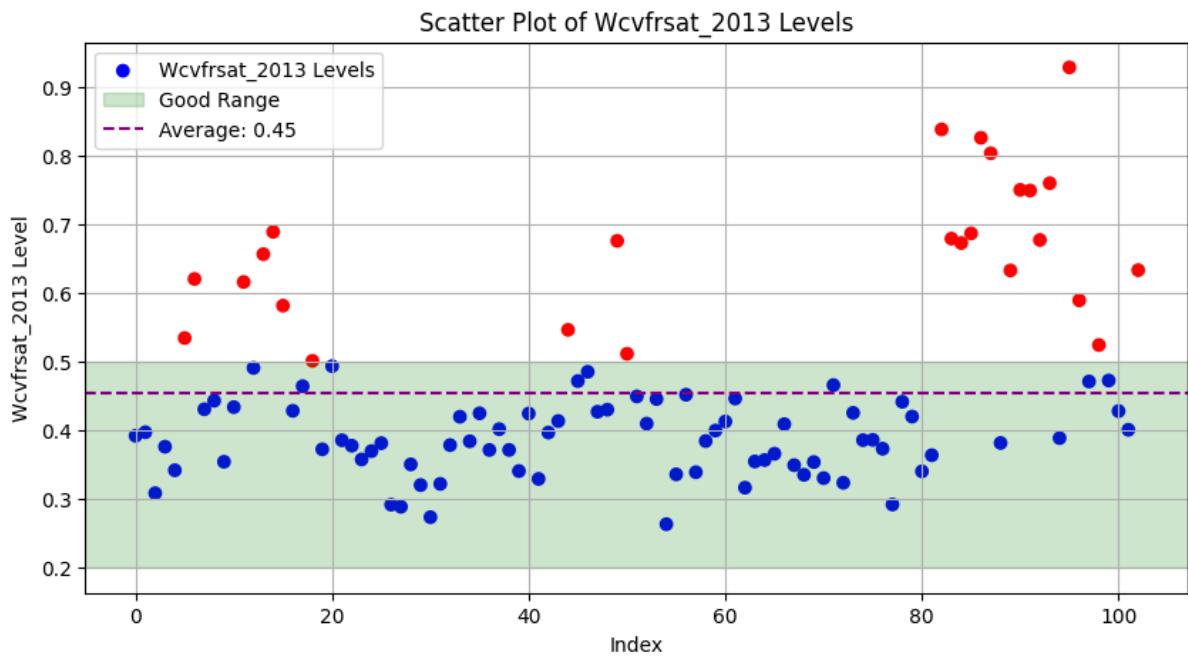
รูป 63 แสดงค่าของข้อมูล Psa asilt ที่มีการระบุช่วงที่เป็นผลดีต่อ din



รูป 64 แสดงค่าของข้อมูล Total carbon ที่มีการระบุช่วงที่เป็นผลดีต่อ din



รูป 65 แสดงค่าของข้อมูล Total nitrogen ที่มีการระบุช่วงที่เป็นผลดีต่อ din



รูป 66 แสดงค่าของข้อมูล Wcvfrsat ที่มีการระบุช่วงที่เป็นผลดีต่อ din

บทที่ 4 ผลการทดสอบและอภิปราย

4.1 รายละเอียดการทดลอง

4.1.1 วัตถุประสงค์ของการทดลอง

- เพื่อตรวจสอบความแม่นยำของโมเดล Machine Learning ในการทำนายแนวโน้มระดับจำนวนองค์ประกอบด้ชนีชี้วัดคุณภาพในดิน
- เพื่อวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต้น (ข้อมูลดิน, สภาพอากาศ, และสถานที่) และตัวแปรตาม (ระดับจำนวนองค์ประกอบด้ชนีชี้วัดคุณภาพของสารเคมีในดิน)

4.1.2 ตัวแปรในการทดลอง

- ตัวแปรต้น
 - ข้อมูลดิน: ค่า pH, ปริมาณแร่ธาตุในดิน (ไนโตรเจน, พอสฟอรัส, โพแทสเซียม ฯลฯ)
 - ข้อมูลสภาพอากาศ: ปริมาณน้ำฝน, อุณหภูมิ, ความชื้นสัมพัทธ์
- ตัวแปรตาม
 - ระดับของการประเมินตั้งแต่ชี้วัดของสารเคมีในดิน 3 ระดับ (มากเกินเกณฑ์ ดีตามเกณฑ์ น้อยเกินเกณฑ์)
- ตัวแปรควบคุม
 - ช่วงเวลาในการเก็บข้อมูล
 - พื้นที่ในการเก็บข้อมูล

4.1.3 โมเดล Machine Learning ที่ใช้

- โมเดลที่ใช้: Random Forest, Support Vector Regression (SVM), Decision Tree, Gradient Boosting, Extreme Gradient Boosting และ Neural Networks
- Dataset: ใช้ข้อมูลดินและสภาพแวดล้อมจากแหล่งข้อมูลสาธารณะ ISRIC-Africa Soil Profiles Database (AfSP)

4.1.4 การออกแบบการทดลอง

- แบ่งข้อมูลเป็นสองชุดโดยลักษณะการแบ่งข้อมูลจะอ้างอิงจาก รูปที่ 41
 - ชุดฝึกสอนโมเดล (Training Set): 80% ของข้อมูล
 - ชุดทดสอบโมเดล (Testing Set): 20% ของข้อมูล
- ใช้วิธี Cross-Validation แบบ k-fold ($k=5$) เพื่อประเมินความแม่นยำของโมเดล

4.2 การทดลอง

4.2.1 ขั้นตอนการทดลอง

1. การเตรียมข้อมูล
 - ทำความสะอาดข้อมูล (Data Cleaning) เช่น จัดการค่าที่ขาดหายไป (Missing Values) การจัดการค่าผิดปกติ (Outlier Treatment) และ การใช้ Feature Engineering ในการสร้างความสัมพันธ์แก่ข้อมูลชุดที่กำหนด
 - ตรวจสอบความสมดุลของกลุ่มตัวอย่าง (Class Imbalance)
2. การสร้างโมเดล
 - ใช้โมเดล Machine Learning ทั้ง 6 รูปแบบ (Random Forest, Support Vector Regression (SVM), Decision Tree, Gradient Boosting, Extreme Gradient Boosting และ Neural Networks)
 - เพื่อทำนายสารเคมีที่สำคัญต่างๆ ในดิน
 - ตั้งค่าพารามิเตอร์เบื้องต้นและปรับแต่งค่าพารามิเตอร์ด้วย Grid Search
3. การเก็บข้อมูลผลลัพธ์
 - เก็บค่าผลลัพธ์ที่ได้จากการทดสอบโมเดล เช่น ค่า R², RMSE, MAE, MAPE
 - วิเคราะห์ผลการทำนายของโมเดลโดยเปรียบเทียบค่าจริงและค่าทำนาย เช่น Percentage Error, Absolute Error เป็นต้น

4.2.2 การแสดงผลการทดลอง

- แสดงผลการทดลองในรูปแบบกราฟ เช่น
 - กราฟเปรียบเทียบค่าความแม่นยำ
 - กราฟเทียบค่าจริงและค่าทำนาย

4.3 การวิเคราะห์ผล

4.3.1 การวิเคราะห์ประสิทธิภาพของโมเดล

เปรียบเทียบประสิทธิภาพของโมเดลทั้ง 6 แบบ โดยพิจารณาค่าความคลาดและค่าความแม่นยำ R2

- Random Forest ให้ผลลัพธ์ดีที่สุด (Percentage Error < 6%)
- SVR มีความแม่นยำต่ำที่สุดในกลุ่ม (Percentage Error > 20%)

4.3.2 การวิเคราะห์ความสัมพันธ์ของตัวแปร

วิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตาม เช่น

- พบร่วมกับ pH และปริมาณแร่ธาตุในดิน (ไนโตรเจนและฟอสฟอรัส) มีความสัมพันธ์เชิงบวกกับระดับคุณภาพของดิน
- ปริมาณน้ำฝนในช่วงเวลาเก็บข้อมูลส่งผลต่อค่าคุณภาพดินในพื้นที่บางแห่ง

4.3.3 ประเด็นที่ค้นพบจากการทดลอง

- พบร่วมกับข้อมูลที่มีตัวแปรตันไม่ครบถ้วน (Missing Data) ส่งผลต่อความแม่นยำของโมเดล
- ความไม่คงที่ของการกระจายตัวข้อมูลแต่ละปีทำให้เห็นแนวโน้มได้ยากและทำให้ความแม่นยำลดลง
- Random Forest เหมาะสมที่สุดสำหรับการใช้งานในโครงการนี้ เนื่องจากให้ค่าความแม่นยำและความน่าเชื่อถือสูง

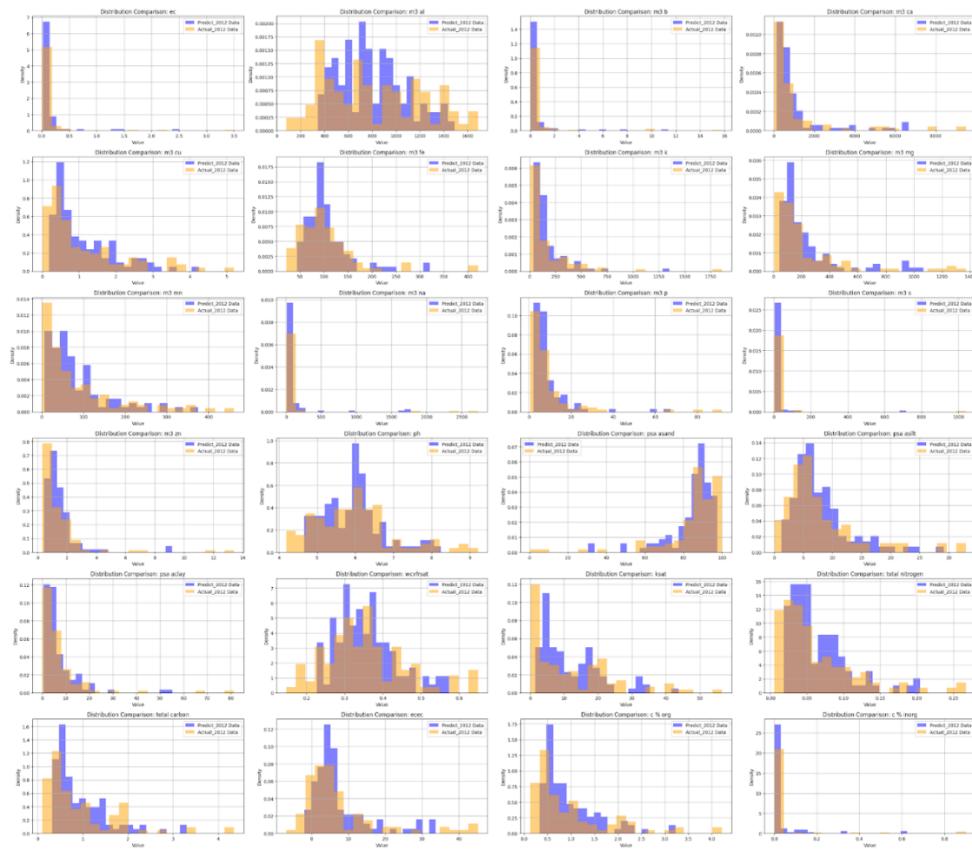
4.3.4 เปรียบเทียบระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012

หลังจากทำการทดสอบระหว่างใช้ค่าที่มี Outlier หรือใช้ค่าที่ได้จากการทำ Outlier treatment ได้ผลสรุปว่า ค่าที่ทำ Outlier treatment นั้นดีกว่าและยังทำให้โมเดลทำนายได้ดีขึ้นด้วย

จากนั้นนำค่าที่ทำ Outlier treatment แล้วไปใส่ในแต่ละโมเดลและนำผลการทำนายที่ได้คำนึงไปประเมินเพื่อตรวจสอบความแม่นยำระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012

จากการตรวจสอบพบว่าค่าที่ทำ Outlier treatment ทำให้โมเดลมีความแม่นยามากขึ้นเทียบกับค่าจริงปี 2012 ดังภาพด้านล่างในแต่ละโมเดลโดยที่ในภาพ Bar plot สีฟ้าคือค่าที่ทำนายได้และสีเหลืองคือค่าที่เป็นค่าจริงจากปี 2012 และพบว่าโมเดลที่มีความแม่นยามากสุด คือ Random forest โดยสังเกตจากค่า Percentage error ที่ค่อนข้างต่ำและมีความสมเหตุผลมากจากทุกโมเดล

1. Random Forest

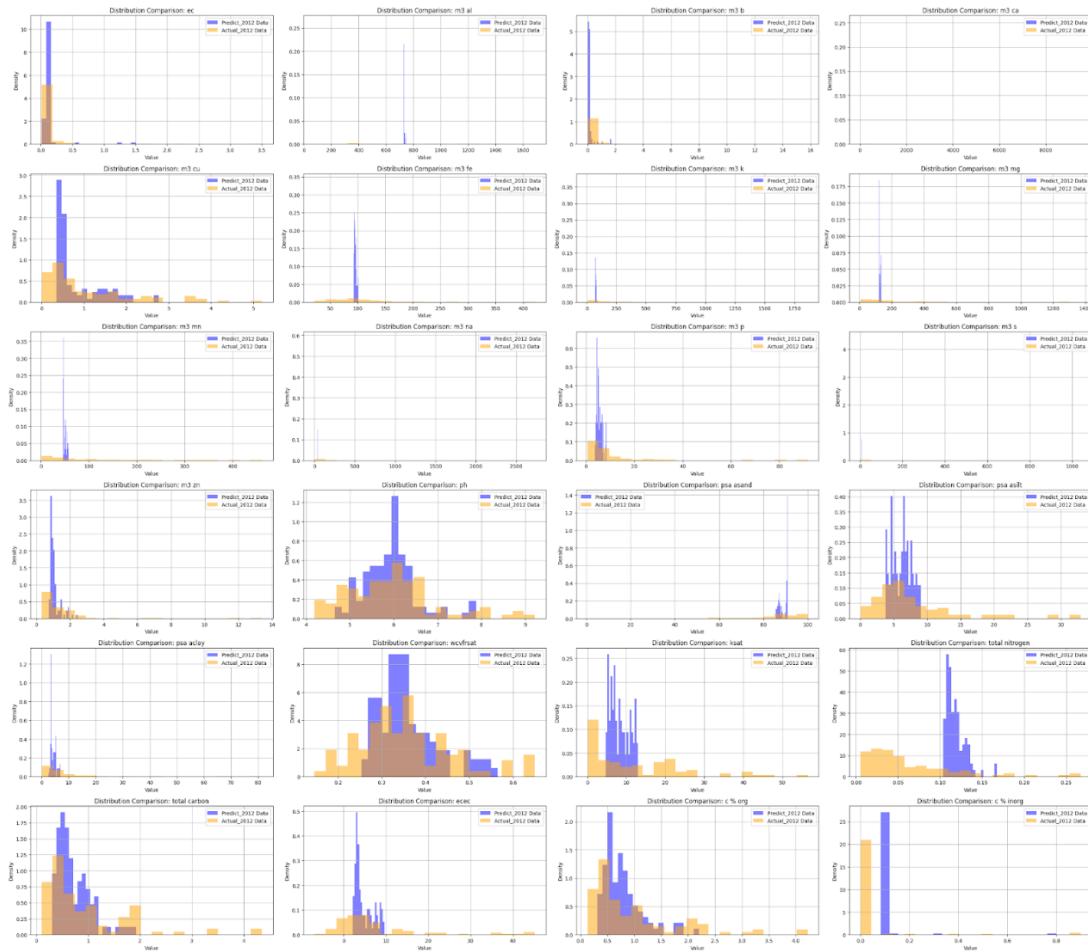


รูป 67 Bar Plot เปรียบเทียบระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (random forest)

Feature	Predict_2012 Mean	Actual_2012 Mean	Correlation	Absolute Error	Percentage Error (%)
0 ec	0.136024	0.137563	0.957528	0.00153942	1.1196
1 m3 al	807.344	805.03	0.82404	2.31418	0.287466
2 m3 b	0.556408	0.558776	0.952597	0.00236778	0.423744
3 m3 ca	1119.9	1134.22	0.95984	14.3207	1.2626
4 m3 cu	1.1147	1.08102	0.959429	0.0336724	3.11486
5 m3 fe	105.65	104.644	0.945598	1.006	0.961353
6 m3 k	172.668	173.117	0.974824	0.448927	0.25932
7 m3 mg	229.827	223.892	0.961017	5.93471	2.6507
8 m3 mn	84.8446	84.6157	0.958233	0.22886	0.27047
9 m3 na	114.628	108.14	0.931921	6.48728	5.99894
10 m3 p	9.07417	9.27083	0.970266	0.196663	2.12131
11 m3 s	20.6449	22.4163	0.994561	1.77143	7.9024
12 m3 zn	1.51578	1.48264	0.953648	0.0331419	2.23534
13 ph	5.95754	5.97163	0.970032	0.0140944	0.236022
14 psa asand	84.6177	84.7983	0.944702	0.180593	0.212967
15 psa asilt	8.02807	7.9094	0.95246	0.118667	1.50033
16 psa aclay	7.25136	7.29221	0.953534	0.0408492	0.560176
17 wcvfrsat	0.356193	0.356641	0.976164	0.00044778	0.125555
18 ksat	11.8461	11.9118	0.970169	0.065686	0.551438
19 total nitrogen	0.0645223	0.0635111	0.965989	0.00101119	1.59215
20 total carbon	0.957335	0.933982	0.945851	0.0233522	2.50028
21 ecec	7.56396	7.42762	0.962227	0.136342	1.83561
22 c % org	0.985027	0.966893	0.956249	0.018134	1.87549
23 c % inorg	0.0323398	0.0291262	0.948989	0.00321359	11.0333

รูป 68 การเปรียบเทียบความแม่นยำระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (random forest)

2. Support Vector Regression (SVR)

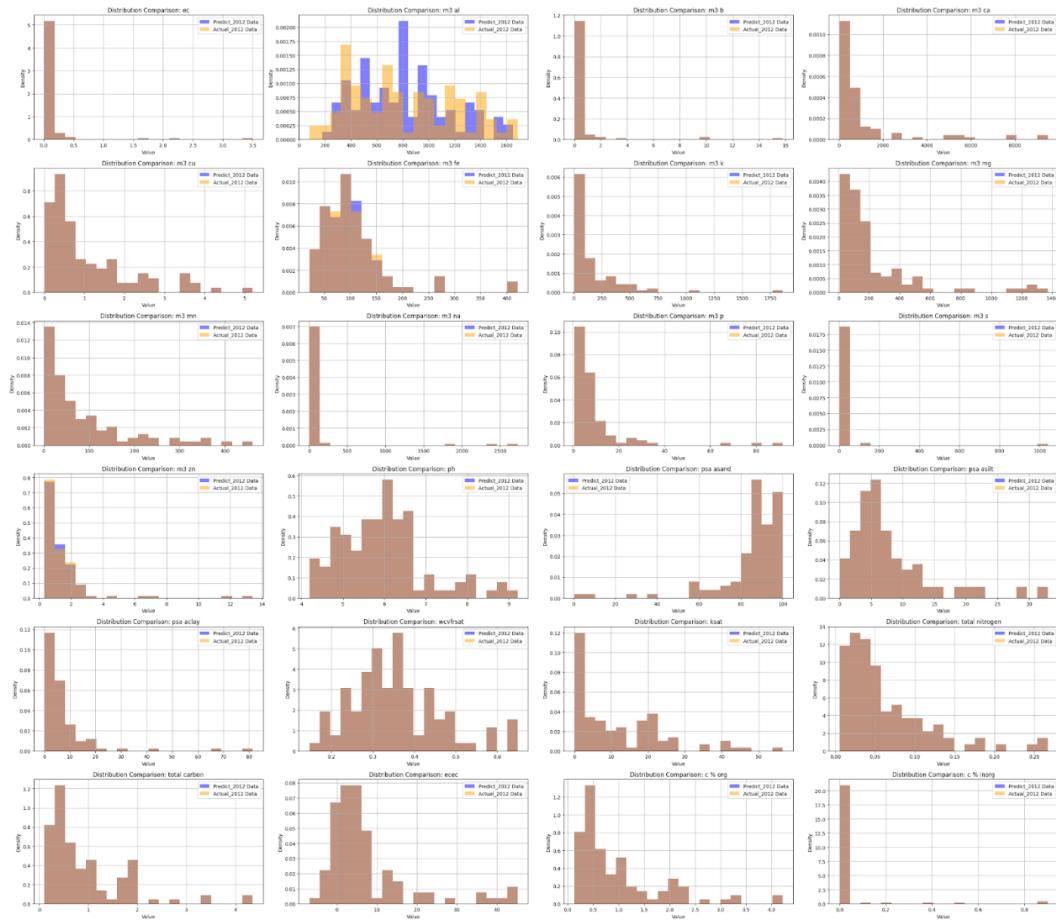


รูป 69 Bar Plot เปรียบเทียบระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (SVR)

Feature	Predict_2012 Mean	Actual_2012 Mean	Correlation	Absolute Error	Percentage Error (%)
0 ec	0.1344	0.137563	0.92112	0.00316269	2.29989
1 m3 al	735.55	805.03	0.468211	69.4796	8.63069
2 m3 b	0.166837	0.558776	0.869211	0.391939	70.1425
3 m3 ca	422.244	1134.22	0.433774	711.979	62.7724
4 m3 cu	0.831875	1.08102	0.691063	0.249148	23.8474
5 m3 fe	95.7843	104.644	0.618731	8.85936	8.46622
6 m3 k	75.329	173.117	0.437631	97.7878	56.4866
7 m3 mg	127.658	223.892	0.474325	96.2336	42.9822
8 m3 mn	50.0912	84.6157	0.542065	34.5245	40.8015
9 m3 na	35.8846	108.14	0.388942	72.2558	66.8166
10 m3 p	5.56145	9.27083	0.584995	3.70938	40.0113
11 m3 s	7.68878	22.4163	0.254085	14.7276	65.7001
12 m3 zn	1.07368	1.48264	0.521562	0.408957	27.5831
13 ph	5.93504	5.97163	0.7558493	0.0365865	0.612671
14 psa asand	89.1593	84.7983	0.470838	4.36105	5.14286
15 psa asilt	6.18678	7.9094	0.582894	1.72263	21.7795
16 psa aclay	4.35616	7.29221	0.52362	2.93606	40.2629
17 wcvfrsat	0.360248	0.356641	0.756772	0.00360643	1.01122
18 ksat	8.28663	11.9118	0.564241	3.62514	30.4332
19 total nitrogen	0.118362	0.0635111	0.623194	0.0548507	86.3639
20 total carbon	0.735339	0.933982	0.737574	0.198644	21.2685
21 eec	4.76598	7.42762	0.633406	2.66164	35.8344
22 c % org	0.783902	0.966893	0.757361	0.182991	18.9257
23 c % inorg	0.105319	0.0291262	0.66959	0.0761929	261.595

รูป 70 การเปรียบเทียบความแม่นยำระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (SVR)

3. Decision Tree

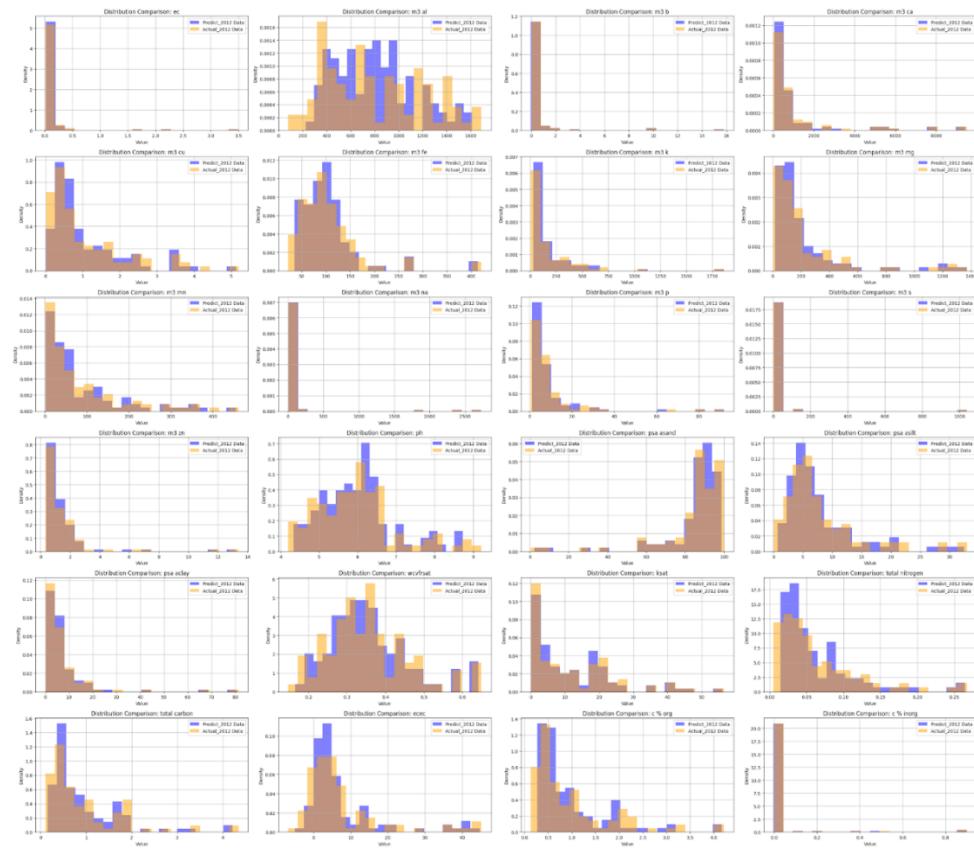


รูป 71 Bar Plot เปรียบเทียบระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (Decision Tree)

Feature	Predict_2012 Mean	Actual_2012 Mean	Correlation	Absolute Error	Percentage Error (%)
0 ec	0.137563	0.137563	1	0	0
1 m3 al	805.03	805.03	0.845876	1.13687e-13	1.41221e-14
2 m3 b	0.558776	0.558776	1	0	0
3 m3 ca	1134.22	1134.22	1	0	0
4 m3 cu	1.08102	1.08102	1	0	0
5 m3 fe	104.644	104.644	0.996395	0	0
6 m3 k	173.117	173.117	1	0	0
7 m3 mg	223.892	223.892	1	0	0
8 m3 mn	84.6157	84.6157	1	0	0
9 m3 na	108.14	108.14	1	0	0
10 m3 p	9.27083	9.27083	1	0	0
11 m3 s	22.4163	22.4163	1	0	0
12 m3 zn	1.48264	1.48264	0.99893	0	0
13 ph	5.97163	5.97163	1	0	0
14 psa asand	84.7983	84.7983	1	0	0
15 psa asilt	7.9094	7.9094	1	0	0
16 psa aclay	7.29221	7.29221	1	0	0
17 wcvfrsat	0.356641	0.356641	1	0	0
18 ksat	11.9118	11.9118	1	0	0
19 total nitrogen	0.0635111	0.0635111	1	0	0
20 total carbon	0.933982	0.933982	1	0	0
21 ecec	7.42762	7.42762	1	0	0
22 c % org	0.966893	0.966893	1	0	0
23 c % inorg	0.0291262	0.0291262	1	0	0

รูป 72 การเปรียบเทียบความแม่นยำระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (Decision Tree)

4. Gradient Boosting

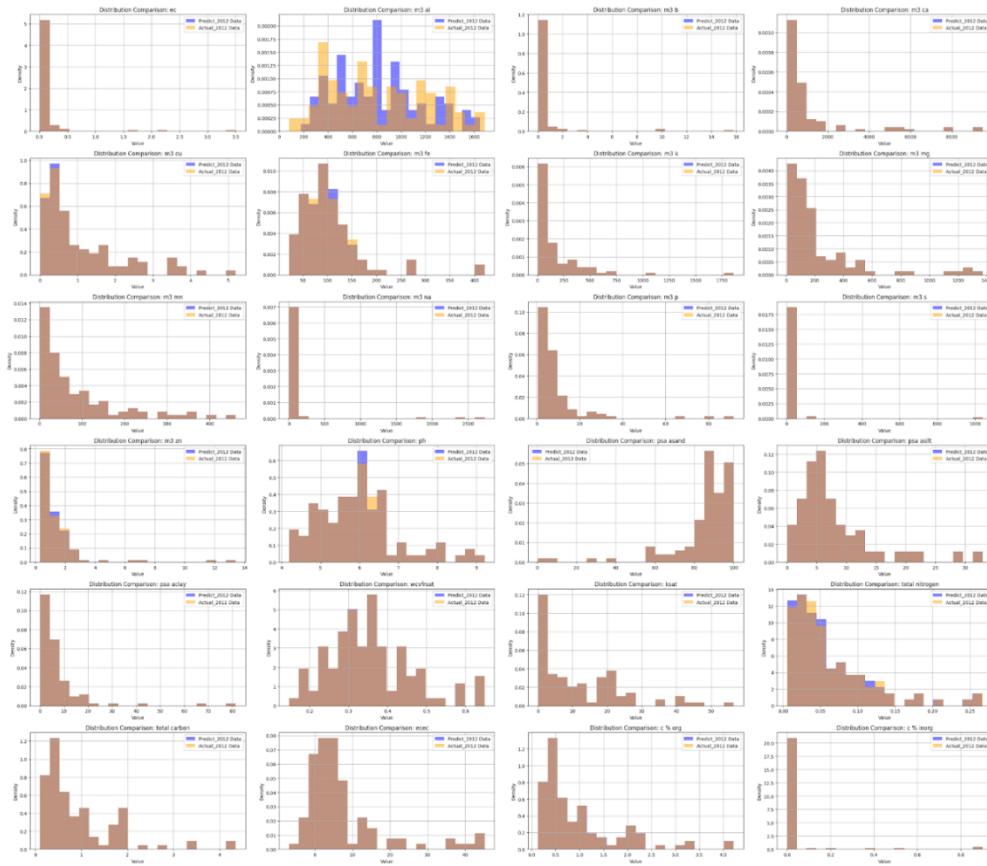


รูป 73 กราฟ Bar Plot เปรียบเทียบระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (Gradient Boosting)

	Feature	Predict_2012 Mean	Actual_2012 Mean	Correlation	Absolute Error	Percentage Error (%)
0	ec	0.137563	0.137563	0.999622	2.77556e-17	2.01766e-14
1	m3_a1	805.03	805.03	0.843465	1.13687e-13	1.41221e-14
2	m3_b	0.558776	0.558776	0.999713	1.11022e-16	1.98688e-14
3	m3_ca	1134.22	1134.22	0.998752	0	0
4	m3_cu	1.08102	1.08102	0.994573	2.22045e-16	2.05402e-14
5	m3_fe	104.644	104.644	0.989508	0	0
6	m3_k	173.117	173.117	0.997224	1.42109e-13	8.20882e-14
7	m3_mg	223.892	223.892	0.996453	5.68434e-14	2.53888e-14
8	m3_mm	84.6157	84.6157	0.996054	1.42109e-14	1.67946e-14
9	m3_na	108.14	108.14	0.999878	5.68434e-14	5.25645e-14
10	m3_p	9.27983	9.27983	0.999141	1.77636e-15	1.91667e-14
11	m3_s	22.4163	22.4163	0.999797	0	0
12	m3_zn	1.48264	1.48264	0.9964	0	0
13	ph	5.97163	5.97163	0.993493	3.55271e-15	5.94932e-14
14	psa_asand	84.7983	84.7983	0.993903	1.42109e-14	1.67584e-14
15	psa_acilt	7.9094	7.9094	0.993894	2.66454e-15	3.36882e-14
16	psa_aclay	7.29221	7.29221	0.997521	8.88178e-16	1.21798e-14
17	wcvfrsat	0.356641	0.356641	0.996768	5.55112e-17	1.5565e-14
18	ksat	11.9118	11.9118	0.997319	0	0
19	total_nitrogen	0.0635111	0.0635111	0.995035	2.77556e-17	4.37019e-14
20	total_carbon	0.933982	0.933982	0.993576	2.22045e-16	2.3774e-14
21	ecec	7.42762	7.42762	0.99655	2.66454e-15	3.58733e-14
22	c_percent_org	0.966893	0.966893	0.99146	1.11022e-16	1.14824e-14
23	c_percent_inorg	0.0291262	0.0291262	0.999946	1.73472e-17	5.95588e-14

รูป 74 การเปรียบเทียบความแม่นยำระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (Gradient Boosting)

5. Extreme Gradient Boosting

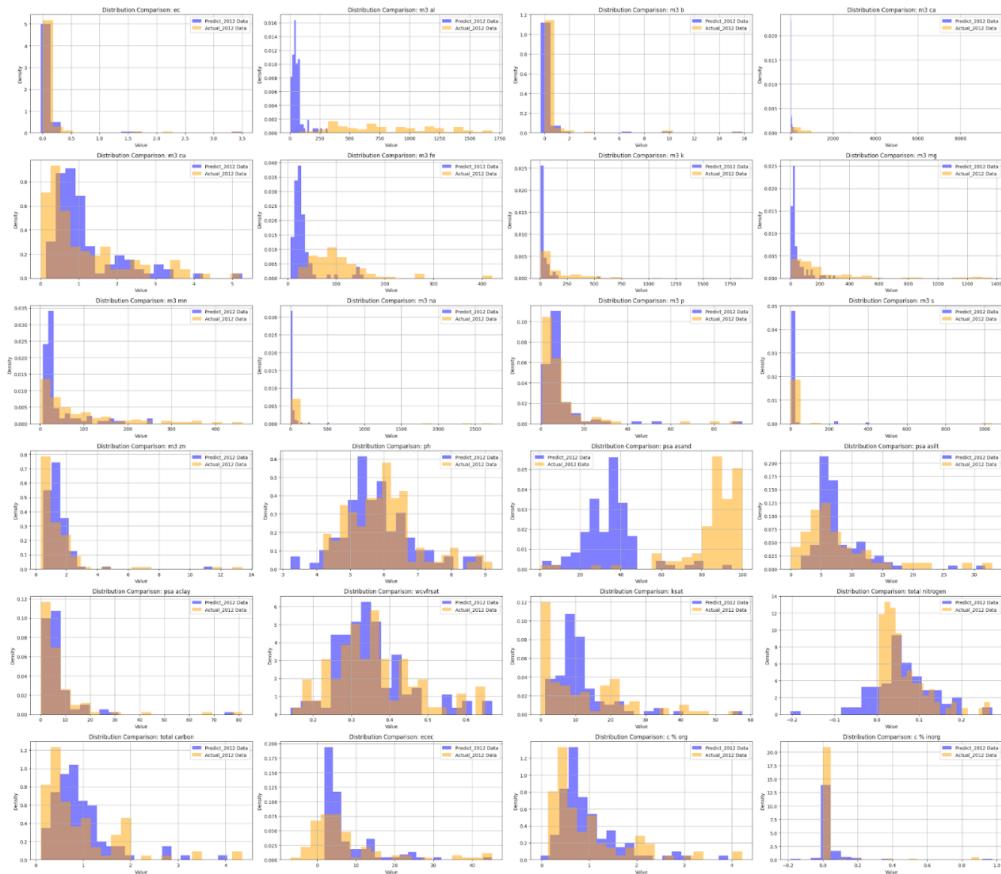


รูป 75 Bar Plot เปรียบเทียบระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (Extreme Gradient Boosting)

Feature	Predict_2012 Mean	Actual_2012 Mean	Correlation	Absolute Error	Percentage Error (%)
0 ec	0.137563	0.137563	0.999999	1.22816e-09	8.92794e-07
1 m3_a1	805.03	805.03	0.845876	1.54369e-05	1.91756e-06
2 m3_b	0.558776	0.558776	1	1.07291e-08	1.92011e-06
3 m3_ca	1134.22	1134.22	1	1.91456e-05	1.68799e-06
4 m3_cu	1.08102	1.08102	1	8.38155e-09	7.75335e-07
5 m3_fe	104.644	104.644	0.996395	5.86408e-06	5.60385e-06
6 m3_k	173.117	173.117	1	3.93592e-06	2.27356e-06
7 m3_mg	223.892	223.892	1	4.61165e-06	2.05977e-06
8 m3_mn	84.6157	84.6157	1	5.5534e-07	6.56308e-07
9 m3_na	108.14	108.14	1	2.46269e-06	2.27731e-06
10 m3_p	9.27083	9.27083	1	1.98738e-07	2.14369e-06
11 m3_s	22.4164	22.4163	1	5.1165e-07	2.28249e-06
12 m3_zn	1.48264	1.48264	0.99893	4.36893e-08	2.94673e-06
13 ph	5.97163	5.97163	1	2.27184e-07	3.8044e-06
14 psa_asand	84.7983	84.7983	1	5.2866e-06	6.23433e-06
15 psa_asilt	7.9094	7.9094	1	1.83626e-07	2.32162e-06
16 psa_aclay	7.29221	7.29221	1	1.18309e-07	1.62241e-06
17 wcvfrsat	0.356641	0.356641	0.999992	2.2233e-08	6.234e-06
18 ksat	11.9118	11.9118	1	1.1511e-07	9.66353e-07
19 total_nitrogen	0.0635111	0.0635111	0.999939	1.97573e-09	3.11084e-06
20 total_carbon	0.933982	0.933982	0.999999	1.79806e-08	1.92515e-06
21 ecec	7.42762	7.42762	1	1.76417e-07	2.37515e-06
22 c_percent_org	0.966893	0.966893	1	2.70874e-08	2.80149e-06
23 c_percent_inorg	0.0291262	0.0291262	0.999997	3.17738e-10	1.0909e-06

รูป 76 การเปรียบเทียบความแม่นยำระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012
(Extreme Gradient Boosting)

6. Neural Networks



รูป 77 Bar Plot เปรียบเทียบระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (Neural Network)

Feature	Predict_2012 Mean	Actual_2012 Mean	Correlation	Absolute Error	Percentage Error (%)
0 ec	0.137629	0.137563	0.970726	6.58784e-05	0.0478896
1 m3 al	60.4236	805.03	0.394016	744.696	92.4942
2 m3 b	0.559386	0.558776	0.966813	0.000610023	0.109171
3 m3 ca	41.184	1134.22	0.625632	1093.04	96.369
4 m3 cu	1.1684	1.08102	0.606796	0.0873754	8.08265
5 m3 fe	31.7404	184.644	0.55492	72.9933	69.6681
6 m3 k	38.6736	173.117	0.757218	134.443	77.6604
7 m3 mg	46.7351	223.892	0.633343	177.157	79.126
8 m3 mn	41.9306	84.6157	0.492958	42.6852	50.4459
9 m3 na	25.6666	108.14	0.690734	82.4738	76.2654
10 m3 p	8.78146	9.27083	0.863352	0.489372	5.27861
11 m3 s	19.3381	22.4163	0.757825	3.07821	13.732
12 m3 zn	1.48154	1.48264	0.677413	0.00109129	0.07356049
13 ph	5.80509	5.97163	0.487821	0.166542	2.78889
14 psa asand	34.7013	84.7983	-0.324189	50.097	59.0778
15 psa asilt	7.87077	7.9094	0.723533	0.0386358	0.48848
16 psa aclay	6.93749	7.29221	0.770584	0.354727	4.86447
17 wcvfrsat	0.352936	0.356641	0.694057	0.003705048	1.03899
18 ksat	11.2397	11.9118	0.66944	0.672669	5.64206
19 total nitrogen	0.0631347	0.0635111	0.559639	0.000376424	0.592691
20 total carbon	0.935418	0.9393892	0.866634	0.00143543	0.153689
21 eec	7.44415	7.42762	0.737757	0.0165322	0.222578
22 c % org	0.973121	0.966893	0.88624	0.006282828	0.644153
23 c % inorg	0.0294161	0.0291262	0.725379	0.000289882	0.995262

รูป 78 การเปรียบเทียบความแม่นยำระหว่างค่าที่ทำนายได้กับค่าจริงของปี 2012 (Neural Network)

บทที่ 5 บทสรุปการดำเนินงานและข้อเสนอแนะ

5.1 สรุปการดำเนินงาน

การพัฒนาระบบวิเคราะห์คุณภาพดินด้วย Machine Learning มุ่งเน้นแก้ไขปัญหาดินเสื่อมสภาพ โดยใช้ข้อมูลคุณภาพดินที่สำคัญ เช่น ค่าความเป็นกรด-ด่าง (pH), ความชื้นในดิน และปริมาณสารอาหารในดิน จากนั้นนำข้อมูลผ่านกระบวนการการทำความสะอาดและเตรียมความพร้อมก่อนสร้างโมเดล Machine Learning

โดยมีการใช้โมเดลทั้งหมด 6 ประเภท ได้แก่ Random Forest, Support Vector Regression (SVM), Decision Tree, Gradient Boosting, Extreme Gradient Boosting และ Neural Networks เพื่อนำมาทดสอบประสิทธิภาพ และเปรียบเทียบค่า R², MAE, RMSE เพื่อศึกษาและเปรียบเทียบประสิทธิภาพของแต่ละโมเดล

ขั้นตอนการดำเนินงานครอบคลุมตั้งแต่การเตรียมข้อมูล การพัฒนาและปรับจูนโมเดล ไปจนถึงการประเมินผลลัพธ์ โดยมีเป้าหมายเพื่อให้โมเดลต้องมีค่าความคลาดเคลื่อนไม่เกิน 6 % เพื่อให้สามารถนำไปต่อยอดเพื่อเชื่อมต่อกับระบบ IoT และพัฒนาสู่การใช้งานจริงในอนาคต เพื่อช่วยเกษตรกรในการจัดการดินอย่างยั่งยืน และลดผลกระทบจากการเปลี่ยนแปลงสภาพภูมิอากาศ

5.2 ความสามารถและข้อจำกัดของ Prototype

ความสามารถของ Prototype

1. ความสามารถในการคาดการณ์คุณภาพดิน

โมเดล Machine Learning สามารถวิเคราะห์และคาดการณ์ระดับจำนวนดัชนีชี้วัดคุณภาพดินได้โดยใช้ข้อมูลพารามิเตอร์สำคัญ เช่น ค่าความเป็นกรด-ด่าง ความชื้น และปริมาณธาตุอาหารในดิน

2. รองรับข้อมูลหลากหลาย

ระบบสามารถจัดการกับข้อมูลหลากหลายรูปแบบ ทั้งตัวเลขและข้อมูลที่ต้องการการปรับขนาด (Standardization)

3. การเปรียบเทียบโมเดลหลายชนิด

4. โมเดลที่พัฒนาประกอบด้วย Random Forest, SVR, Decision Tree, Gradient Boosting, Extreme Gradient Boosting และ Neural Network ซึ่งช่วยให้สามารถเลือกโมเดลที่มีประสิทธิภาพดีที่สุดได้

5. ความยืดหยุ่นในการปรับปรุง

สามารถเพิ่มข้อมูลใหม่หรือปรับจูนพารามิเตอร์ของโมเดลเพื่อปรับปรุงความแม่นยำและประสิทธิภาพได้

6. มีประโยชน์ต่อการตัดสินใจ

ช่วยให้เกษตรกรและนักวิจัยมีข้อมูลในการตัดสินใจปรับปรุงดินหรือบริหารจัดการทรัพยากรในพื้นที่ การเกษตร

ข้อจำกัดของ Prototype

1. คุณภาพของข้อมูล

ผลลัพธ์ของโมเดลขึ้นอยู่กับความครบถ้วนและความถูกต้องของข้อมูล หากข้อมูลขาดหายหรือมีความผิดพลาด อาจส่งผลกระทบต่อความแม่นยำของการคาดการณ์

2. ไม่ครอบคลุมทุกปัจจัยที่ส่งผลต่อดิน

ข้อมูลที่ใช้จึงไม่ได้ครอบคลุมปัจจัยด้านสิ่งแวดล้อมทั้งหมด เช่น สภาพภูมิอากาศ หรือการชะล้างหน้าดินจากการไหลของน้ำ

3. ข้อจำกัดในการใช้งานจริง

การนำต้นแบบไปใช้งานจริงอาจต้องปรับแต่งเพิ่มเติมให้เหมาะสมกับแต่ละพื้นที่ เนื่องจากลักษณะของดินและพืชพันธุ์แตกต่างกัน

4. การวิเคราะห์ผลกราฟบรรยายภาพ

Prototype ยังไม่สามารถคาดการณ์ผลกราฟบรรยายภาพได้ เช่น การพื้นฟูคุณภาพดิน หรือความเปลี่ยนแปลงจาก Climate Change ได้อย่างแม่นยำ

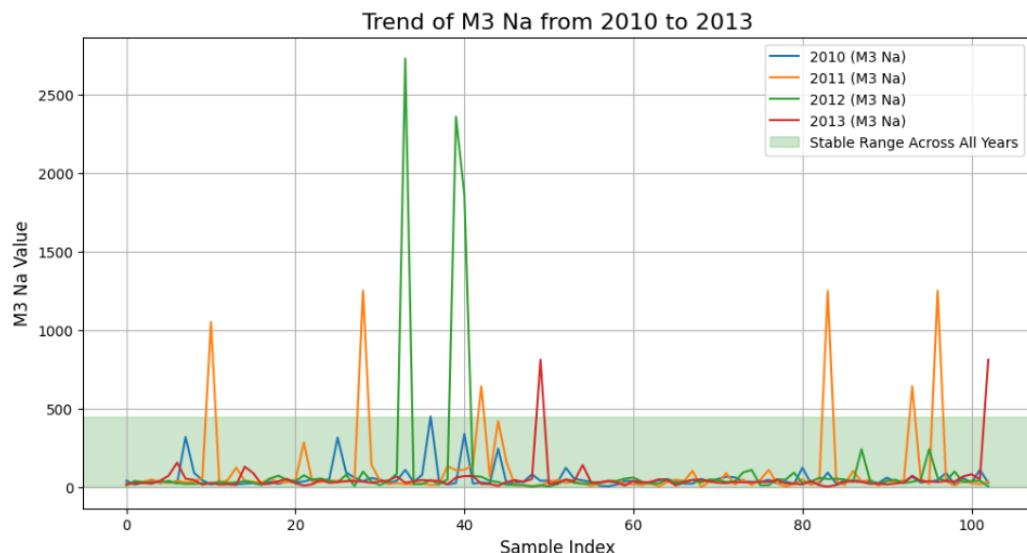
5. ข้อจำกัดด้านทรัพยากร

การประมวลผลโมเดลต้องใช้ทรัพยากรคอมพิวเตอร์ในระดับหนึ่ง ซึ่งอาจไม่เหมาะสมสำหรับพื้นที่ที่ห่างไกล หรือผู้ใช้งานที่มีข้อจำกัดด้านเทคโนโลยี

5.3 ปัญหาที่พบในการดำเนินงาน

ในระหว่างการดำเนินงานพบปัญหาหลายประการที่ส่งผลต่อความแม่นยำและประสิทธิภาพของการทำนาย ดังนี้

1. แนวโน้มไม่คงที่: ข้อมูลที่นำมาใช้ในการทำนายมีแนวโน้มที่ไม่สม่ำเสมอ ทำให้ยากต่อการสร้างโมเดลที่สามารถคาดการณ์ได้อย่างแม่นยำ



รูป 79 แนวโน้มของ M3 Al ของทั้ง 4 ปี

2. จำนวนปีในการหาแนวโน้มไม่เพียงพอ: ระยะเวลาที่ใช้ในการเก็บข้อมูลไม่เพียงพอสำหรับการวิเคราะห์แนวโน้มในระยะยาว ส่งผลให้โมเดลไม่สามารถจับแนวโน้มที่แท้จริงได้
3. ปริมาณข้อมูลน้อย: ข้อมูลที่ใช้ในการฝึกโมเดลมีปริมาณน้อย ทำให้ไม่สามารถหาแนวโน้มแบบ Non-Linear ได้อย่างมีประสิทธิภาพ
4. ความแม่นยำในการประเมินคุณภาพสารเคมีในดิน: การประเมินคุณภาพสารเคมีในดินต้องอิงจากไฟล์ผลลัพธ์ที่มีความแม่นยำสูง ซึ่งอาจทำให้เกิดข้อจำกัดในการวิเคราะห์

5.4 แนวทางการพัฒนาในอนาคต

1. การเพิ่มความแม่นยำของโมเดล Machine Learning
 - รวบรวมข้อมูลเพิ่มเติมจากพื้นที่การเกษตรหลากหลายภูมิภาค เพื่อเพิ่มความหลากหลายชุดข้อมูล

- ทดลองใช้เทคนิค Machine Learning ขั้นสูง เช่น Deep Learning หรือ Ensemble Learning เพื่อเพิ่มประสิทธิภาพการทำนาย
- ปรับจุนพารามิเตอร์ (Hyperparameter Tuning) ให้เหมาะสมยิ่งขึ้น

2. การบูรณาการระบบ IoT

- พัฒนาระบบเซ็นเซอร์ IoT ที่สามารถเก็บข้อมูลติดแบบเรียลไทม์ เช่น ความชื้น, อุณหภูมิ, และค่าความเป็นกรด-ด่าง
- เข้ามต่อระบบเซ็นเซอร์กับโมเดล Machine Learning เพื่อให้ผลการวิเคราะห์แบบทันทีและแม่นยำ

3. การพัฒนาส่วนติดต่อผู้ใช้งาน (User Interface)

- สร้างแพลตฟอร์มที่ใช้งานง่าย เช่น แอปพลิเคชันมือถือหรือเว็บไซต์ เพื่อให้เกษตรกรสามารถเข้าถึงข้อมูลและคำแนะนำได้สะดวก
- แสดงผลลัพธ์ในรูปแบบที่เข้าใจง่าย เช่น แผนภาพ, กราฟ, หรือคำแนะนำที่ปรับตามพื้นที่

4. การสนับสนุนการตัดสินใจของเกษตรกร

- พัฒนาโมเดลที่สามารถให้คำแนะนำสำหรับการปรับปรุงคุณภาพดิน เช่น การเลือกใช้ปุ๋ย การปลูกพืชหมุนเวียน หรือการจัดการน้ำ
- เพิ่มฟังก์ชันสำหรับการคาดการณ์ผลผลิตหรือรายได้ เพื่อช่วยวางแผนการทำเกษตรกรรม

5. การติดตามและปรับปรุงอย่างต่อเนื่อง

- เก็บข้อมูลการใช้งานจริงจากผู้ใช้งาน เพื่อนำมาปรับปรุงประสิทธิภาพของระบบ
- เพิ่มฟีเจอร์ใหม่ตามความต้องการของผู้ใช้งาน เช่น การแจ้งเตือนปัญหา หรือคำแนะนำรายสัปดาห์

อ้างอิง

- รายงานจากองค์การอาหารและการเกษตรแห่งสหประชาชาติ (FAO): Status of the World's Soil Resources (SWSR)
- Ecowatch. (n.d.). Land Degradation: UN Report. สืบค้นจาก <https://www.ecowatch.com/land-degradation-un-report.html>
- Food and Agriculture Organization of the United Nations. (2023). Soil Degradation and Restoration. สืบค้นจาก <https://www.fao.org/soils-portal/soil-degradation-restoration/en/>
- Oldeman, L. R. (1994). The Global Extent of Soil Degradation. World Soil Resources Reports.Sinha, A. et al. (2021). Soil Quality Prediction Using Machine Learning Algorithms. International Journal of Agricultural and Environmental Research.
- Our World in Data. (n.d.). Share of Degraded Land. สืบค้นจาก <https://ourworldindata.org/grapher/share-degraded-land>
- The Climate Fresk. (n.d.). The Climate Fresk Cards. สืบค้นจาก <https://restore.fresqueduclimat.org/memo/en-US/game/list/>
- The Guardian. (2022). United Nations: 40% of Planet's Land Degraded. สืบค้นจาก <https://www.theguardian.com/environment/2022/apr/27/united-nations-40-per-cent-planet-land-degraded>