



เรื่อง

การพยากรณ์อุณหภูมิเฉลี่ยของผิวโลก

จัดทำโดย

นางสาว ปรียานุช	สุภาสীব	รหัสนักศึกษา 61070306
นางสาว อิงฟ้า	ภูติวรรณถ	รหัสนักศึกษา 61070330

เสนอ

ผศ.ดร.กนกวรรณ อัจฉริยะชาญวณิช

รายงานนี้เป็นส่วนหนึ่งของวิชา Introduction to Business Intelligence (06026115)

สาขาวิทยาการข้อมูลและการวิเคราะห์เชิงธุรกิจ คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ประจำภาคเรียนที่ 2 ปีการศึกษา

คำนำ

รายงานฉบับนี้เป็นส่วนหนึ่งของรายวิชา Introduction to Business Intelligence (06026115) โดยมีจุดประสงค์เพื่อการศึกษาความรู้ที่ได้จากการใช้ Alteryx ในการทำ Data Mining ผู้จัดทำได้เลือกหัวข้อการพยากรณ์อุณหภูมิเฉลี่ยของผิวโลกในการจัดทำรายงาน

ผู้จัดทำขอขอบพระคุณ ผศ.ดร.กนกวรรณ อัจฉริยะชาญวิช ผู้ให้ความรู้และแนวทางการศึกษา ผู้จัดทำหวังว่ารายงานฉบับนี้จะให้ความรู้และเป็นประโยชน์แก่ผู้อ่าน หากมีข้อผิดพลาดประการใด ผู้จัดทำขอน้อมรับไว้และขออภัยมา ณ ที่นี้ด้วย

ผู้จัดทำ

สารบัญ

	หน้า
ก คำนำ	ก
ข สารบัญ	ข
1 บทสรุปผู้บริหาร	1
2 การอธิบายปัญหา	2
3 การอธิบายข้อมูล	2
4 การเตรียมข้อมูล	3
5 วิธีการไขปัญหาด้วย Data Mining	11
6 บทสรุป	19
7 บรรณานุกรม	20

1.บทสรุปบริหาร

ในปัจจุบันสภาวะภูมิอากาศของโลกมีการเปลี่ยนแปลงไปอย่างมาก เนื่องมาจากสาเหตุภาวะโลกร้อน (Global Warming) โดยสังเกตได้จากการที่อุณหภูมิของโลกที่มีค่าสูงขึ้นเรื่อยๆในทุกๆปี ซึ่งสาเหตุหลักของปัญหานี้เกิดจากก๊าซเรือนกระจก (Green House Gas) กล่าวคือมลภาวะจากก๊าซที่มีคุณสมบัติในการดูดซับพลังงานความร้อนที่สะท้อนจากโลกกลับสู่ชั้นบรรยากาศ เช่น คาร์บอนไดออกไซด์ มีเทน ไนตรัสออกไซด์ คลอโรฟลูออโรคาร์บอน เป็นต้น ก๊าซเหล่านี้ใช้ในกระบวนการผลิตของโรงงานอุตสาหกรรม บ้างเกิดจากไอเสียจากการเผาไหม้เชื้อเพลิงในภาคขนส่ง หรือเมื่อภาคอุตสาหกรรมการผลิตไฟฟ้า หรือ การกระทำใดๆ ก๊าซดังกล่าวมีความสามารถในการกักเก็บความร้อนบางส่วนไว้ในโลก โดยไม่ปล่อยความร้อนสู่บรรยากาศ ด้วยลักษณะที่คล้ายกับหลักการของเรือนกระจกที่ใช้ในการปลูกพืช จึงมักเรียกกันว่า ปรากฏการณ์เรือนกระจก (Greenhouse Effect)

ผลจากการที่อุณหภูมิเฉลี่ยของโลกสูงขึ้นทำให้เกิดการเปลี่ยนแปลงของสภาพภูมิอากาศ (Climate Change) จะส่งผลกระทบต่อดำรงอยู่ของสิ่งมีชีวิตทั้งมนุษย์สัตว์โดยตรง และยังส่งผลทางอ้อมต่อความมั่นคงทางสังคม และเศรษฐกิจของประเทศ อาทิเช่น การส่งผลกระทบต่อผลผลิตทางการเกษตร ผลิตภาพของแรงงานซึ่งเกิดจากสภาวะการทำงานที่ไม่เป็นปกติ ทำงานได้ยากลำบากขึ้น สุขภาพทางจิตใจแย่ลงซึ่งจะทำให้อัตราการเติบโตของเศรษฐกิจโลกได้รับผลกระทบตามมา

ผู้จัดทำจึงตระหนัก และเลือกที่จะศึกษาในเรื่องก๊าซเรือนกระจกดังกล่าวโดยใช้เทคนิค Data Mining ด้วยเทคนิคการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression) ในการพยากรณ์ว่าอุณหภูมิเฉลี่ยของโลกจะสูงขึ้น หรือลดลงตามปริมาณก๊าซที่ก่อให้เกิดสภาวะเรือนกระจก เพื่อนำค่าทำนายไปใช้วิเคราะห์ในการลดปริมาณของก๊าซเรือนกระจก และวางแผนที่จะบรรณรงคทางด้านทรัพยากรธรรมชาติในทุกๆด้าน รวมถึงการอนุรักษ์สิ่งแวดล้อมอย่างเหมาะสม ดังนั้นการทำให้ผู้คนได้ตระหนักถึงผลกระทบที่รุนแรงจากภาวะอุณหภูมิโลกที่สูงขึ้น จะทำให้การดำรงชีวิตในปัจจุบันของทุกคนได้คงอยู่ต่อไป การปรับตัว และสร้างภูมิคุ้มกันให้เข้ากับสภาพแวดล้อมที่เปลี่ยนไป เป็นสิ่งที่ควรดำเนินการควบคู่กันไป และต้องดำเนินการโดยเร่งด่วนอย่างมีประสิทธิภาพ

ซึ่งทางผู้จัดทำได้เลือกข้อมูลที่จะทำการวิเคราะห์มาจากทั้ง 2 ตาราง คือ จากตารางที่ 3.1 berkley_earth_GlobalTemperatures ได้เลือกใช้ตัวแปร LandAverageTemperature มาเป็นตัวแปรตามสำหรับสมการทำนาย และตัวแปร dt และจากตารางที่ 3.2 noaa_aggi_forcing ทำการเลือกตัวแปร ก๊าซเรือนกระจกทั้งหมด 6 ตัวแปร คือ CO₂, CH₄, N₂O, CFC12, CFC11, 15_minor มาเป็นตัวแปรต้นสำหรับสมการทำนาย และตัวแปร year อีกหนึ่งตัวแปรในการสร้างสมการพยากรณ์

2.คำอธิบายปัญหา

เป้าหมายของการทำ Data Mining

เป้าหมายของการวิเคราะห์ คือ การพยากรณ์อุณหภูมิเฉลี่ยของโลกว่าจะสูงขึ้น หรือ ลดลงด้วยตัวแปรอิสระ คือ ปริมาณก๊าซที่ก่อให้เกิดภาวะเรือนกระจก โดยทางเราได้เลือกใช้เทคนิคการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression) ในการทำนายค่าเฉลี่ยของอุณหภูมิโลก ซึ่งการที่อุณหภูมิของโลกมีแนวโน้มเริ่มสูงขึ้นในทุกๆปีทำให้ทางผู้จัดทำจึงใช้ค่าทำนายเพื่อนำไปช่วยในการลดปริมาณก๊าซเรือนกระจกที่มีการปล่อยออกมามากเกินไป และช่วยในการวางแผนการจัดการทรัพยากรธรรมชาติในอนาคตได้อย่างเหมาะสม

3.การอธิบายข้อมูล

การเก็บรวบรวมข้อมูล

ข้อมูลที่นำมาใช้เป็นข้อมูลทุติยภูมิ (Secondary Data) เป็นข้อมูลที่ไม่ได้เก็บรวบรวมเอง โดยนำชุดข้อมูลมาจากเว็บไซต์ Data.world เป็นเว็บไซต์ที่เก็บรวบรวมข้อมูลมาให้เรียบร้อยแล้ว และหัวข้อเรื่องที่ใช้คือ Global Temp & Greenhouse Gas

ข้อมูลที่ใช้

ตารางที่ 3.1 berkley_earth_GlobalTemperatures

ตัวแปร	ประเภทข้อมูล	คำอธิบาย
dt	date	ปี-เดือน-วัน
LandAverageTemperature	decimal	ค่าเฉลี่ยอุณหภูมิผิวโลก
LandAverageTemperatureUncertainty	decimal	ความไม่แน่นอนของค่าเฉลี่ยอุณหภูมิผิวโลก
LandMaxTemperature	decimal	ค่าอุณหภูมิผิวโลกที่สูงที่สุด
LandMaxTemperatureUncertainty	decimal	ความไม่แน่นอนของค่าอุณหภูมิผิวโลกที่สูงที่สุด
LandMinTemperature	decimal	ค่าอุณหภูมิผิวโลกที่ต่ำที่สุด
LandMinTemperatureUncertainty	decimal	ความไม่แน่นอนของค่าอุณหภูมิผิวโลกที่ต่ำที่สุด
LandAndOceanAverageTemperature	decimal	อุณหภูมิเฉลี่ยผิวโลกและในมหาสมุทร
LandAndOceanAverageTemperatureUncertainty	decimal	ความไม่แน่นอนของอุณหภูมิเฉลี่ยผิวโลกและในมหาสมุทร

จากตารางที่ 3.1 ได้แสดงข้อมูลรายละเอียดของตัวแปรต่างๆของค่าเฉลี่ยอุณหภูมิผิวโลกทั่วโลกในทุกเดือนตั้งแต่ปี ค.ศ. 1750 จนถึงปี 2015 มีจำนวนทั้งสิ้น 3,192 แถวโดยในการทำพยากรณ์นี้ เราได้เลือกใช้ตัวแปร LandAverageTemperature คือ ค่าเฉลี่ยอุณหภูมิโลกเพื่อเป็นตัวแปรตามในการวิเคราะห์สมการถดถอยเชิงเส้น และตัวแปร dt คือ ปี-เดือน-วัน ที่มีการเก็บข้อมูลค่าเฉลี่ยอุณหภูมิโลก

ตารางที่ 3.2 noaa_aggi_forcing

ตัวแปร	ประเภทข้อมูล	คำอธิบาย
year	string	ปี ค.ศ
CO2	decimal	ก๊าซคาร์บอนไดออกไซด์
CH4	decimal	ก๊าซมีเทน
N2O	decimal	ก๊าซไนตรัสออกไซด์
CFC12	decimal	ก๊าซไดคลอโรไดฟลูโอโรมีเทน
CFC11	decimal	ก๊าซไตรคลอโรฟลูโอโรมีเทน
15_minor	decimal	ก๊าซเรือนกระจก 15-minor
total	decimal	ผลรวม
co2_eq_ppm_total	decimal	ผลรวมคาร์บอนไดออกไซด์เทียบเท่า
aggi_1990_1	decimal	ดัชนีก๊าซเรือนกระจก ใช้ปี 1990 เป็นปี พื้นฐานที่มีค่าเท่ากับ 1
aggi_change	decimal	ดัชนีก๊าซเรือนกระจก

จากตารางที่ 3.2 แสดงข้อมูลรายละเอียดของตัวแปรก๊าซที่ส่งผลทำให้เกิดสภาวะเรือนกระจก รวมถึงดัชนีก๊าซเรือนกระจกซึ่งมีการเก็บข้อมูลปริมาณก๊าซเรือนกระจกมาตั้งแต่ปี ค.ศ. 1979 จนถึงปี 2015 เป็นจำนวนข้อมูลทั้งหมด 39 แถว ทางผู้จัดทำได้เลือกตัวแปรก๊าซเรือนกระจกที่นำไปใช้ในการพยากรณ์ทั้งหมด 6 ตัวแปร คือ CO2, CH4, N2O, CFC12, CFC11, 15_minor และได้เลือกตัวแปร year คือ ปีค.ศ.ที่ได้ทำการเก็บข้อมูลปริมาณก๊าซเรือนกระจก

4.การเตรียมข้อมูล

1.นำ Input Data Tool มาวางใน Workflow จากนั้นเชื่อมต่อข้อมูลที่มีชื่อไฟล์ว่า berkley_earth_Global Temperatures.csv ดังที่แสดงในรูปที่ 4.1

The screenshot shows the 'Connect a File or Database' tool in Alteryx. The file path is 'D:\Pin Engfah\DSBA2-2\BI\project\global\berkley_earth_GlobalTemperatures.csv'. The 'Options' section is configured with 'Record Limit' set to 1, 'File Format' as 'Comma Separated Values', 'Search SubDirs' unchecked, 'Output File Name as Field' set to 'No', 'Delimiters' as ',', and 'First Row Contains Field Names' checked. The 'Preview first 100 records' section shows a table with columns 'dt', 'LandAverageTemperature', and 'LandAverageTemperature'. The 'Results - Input Data (1) - Output' section displays a table with 9 fields and 3,192 records.

Record	dt	LandAverageTemperature	LandAverageTemperature
1	1750-01-01	3.0340000000000003	3.574
2	1750-02-01	3.083	3.702
3	1750-03-01	5.626	3.076
4	1750-04-01	8.49	2.451
5	1750-05-01	11.573	2.072
6	1750-06-01	12.937000000000001	1.724
7	1750-07-01	15.868	1.911
8	1750-08-01	14.75	2.231
9	1750-09-01	11.412999999999998	2.637
10	1750-10-01	6.366999999999999	2.668
11	1750-11-01	[Null]	[Null]

รูปที่ 4.1 การเชื่อมต่อข้อมูลด้วย Input Data Tool

2.นำ Select Tool มาวางต่อจาก Input Data Tool จากนั้นเลือกข้อมูลแค่ Landaveragetemperature และ dt แล้วเปลี่ยน Type ของข้อมูล Landaveragetemperature จาก String เป็น int64 ดังแสดงรูปที่ 4.2

The screenshot shows the 'Select' tool in Alteryx. The 'Options' section shows the 'Field' list with 'dt' and 'LandAverageTemperature' selected. The 'Type' for 'dt' is 'V_String' and for 'LandAverageTemperature' is 'Int64'. The 'Results - Select (2) - Output' section displays a table with 2 fields and 3,192 records.

Record	dt	LandAverageTemperature
1	1750-01-01	3
2	1750-02-01	3
3	1750-03-01	5
4	1750-04-01	8
5	1750-05-01	11
6	1750-06-01	12
7	1750-07-01	15
8	1750-08-01	14
9	1750-09-01	11
10	1750-10-01	6
11	1750-11-01	[Null]

รูปที่ 4.2 การใช้ Select Tool เพื่อเลือกฟิลด์ที่ต้องการ และเปลี่ยน type ข้อมูล

3.จากรูปที่ 4.3 นำ Text To Columns Tool หรือ เครื่องมือที่ใช้ในการตัดค่า และให้ทำการตั้งค่าใน Configuration ดังนี้

- เลือกคอลัมน์ dt ตรงส่วน column to split
- กำหนดส่วน Delimiters คือ “-”
- กำหนดส่วน number of columns(จำนวนคอลัมน์) คือ 3

Select Column to Split

Column to split: dt Delimiters: -

☒ Split to columns

Number of columns: 3

Extra characters: Leave extra in last column

Output root name:

☐ Split to rows

Advanced options

- ☐ Ignore delimiters in quotes
- ☐ Ignore delimiters in single quotes
- ☐ Ignore delimiters in parentheses
- ☐ Ignore delimiters in brackets
- ☐ Skip empty columns

Results - Text To Columns (4) - Output

Record	dt	LandAverageTemperature	1	2	3
1	1750-01-01	3	1750	01	01
2	1750-02-01	3	1750	02	01
3	1750-03-01	5	1750	03	01
4	1750-04-01	8	1750	04	01
5	1750-05-01	11	1750	05	01
6	1750-06-01	12	1750	06	01
7	1750-07-01	15	1750	07	01
8	1750-08-01	14	1750	08	01
9	1750-09-01	11	1750	09	01
10	1750-10-01	6	1750	10	01
11	1750-11-01	[Null]	1750	11	01

รูปที่ 4.3 การนำ Text To Columns Tool มาใช้ในการตัดค่าในคอลัมน์ dt

4.ใช้ Select Tool ในการเลือกฟิลด์ที่ต้องการ คือ 1 และ Landaveragetemperature แล้วเปลี่ยนชื่อฟิลด์ที่ชื่อ 1 เป็น Year ดังรูปที่ 4.4

Options

Field	Type	Size	Rename
dt	V_String	254	
1	V_String	254	Year
2	V_String	254	
3	V_String	254	
LandAvera...	Int64	8	
*Unknown	Unknown	0	

Results - Select (5) - Output

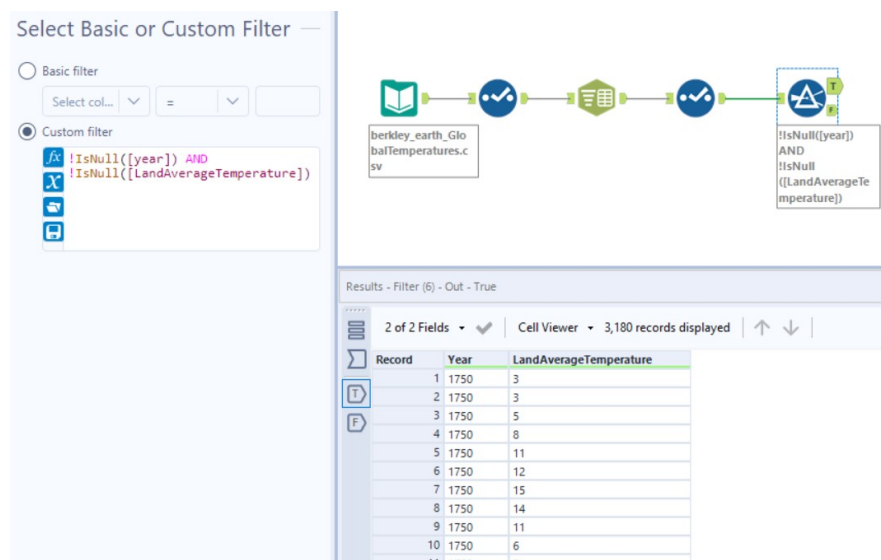
Record	Year	LandAverageTemperature
1	1750	3
2	1750	3
3	1750	5
4	1750	8
5	1750	11
6	1750	12
7	1750	15
8	1750	14
9	1750	11
10	1750	6
11	1750	[Null]

รูปที่ 4.4 การใช้ Select Tool เพื่อเปลี่ยนชื่อฟิลด์ เลือกฟิลด์ที่ต้องการ

5. จากรูปที่ 4.5 เนื่องจากข้อมูลที่นำมาใช้นั้นมีข้อมูลบางส่วนที่เป็นค่า Null จึงต้องมีการกรอง ค่า Null ออก โดยใช้ Filter Tool และทำการตั้งค่า Configuration เพื่อไม่ให้ข้อมูลมีค่า Null โดยเลือก Custom filter จากนั้นใส่คำสั่งเพื่อกรองข้อมูลที่มี Null ดังนี้

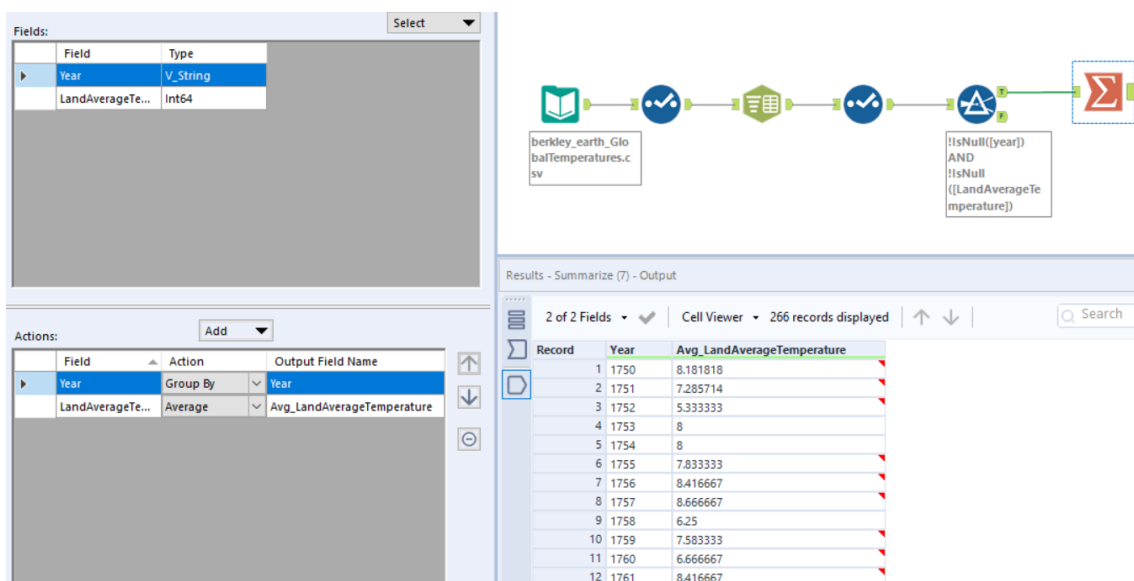
!IsNull([year]) AND

!IsNull([LandAverageTemperature])



รูปที่ 4.5 การใช้ Filter Tool ในการกรองค่า Null

6. ทำการ Summarize ข้อมูล โดยใช้ Summarize Tool ในส่วนของ Configuration ให้ทำการจัดค่า Year แบบ Group by, LandAverageTemperature แบบ Average ตามลำดับในรูปที่ 4.6



รูปที่ 4.6 การใช้ Summarize Tool เพื่อจัดกลุ่มข้อมูล และหาค่าเฉลี่ย

7. นำ Input Data Tool มาวางใน Workflow จากนั้นเชื่อมต่อกับข้อมูล noaa_aggi_forcing.csv ดังที่แสดงในรูปที่ 4.7

Connect a File or Database
D:\Pin English\DSBA2-2\B\project\global\noaa_aggi_forcing.csv

Options

Name	Value
1 Record Limit	
2 File Format	Comma Separated Value (*.csv)
3 Search SubDirs	<input type="checkbox"/>
4 Output File Name as Field	No
5 Delimiters	,
6 First Row Contains Field Names	<input checked="" type="checkbox"/>

Preview (first 100 records)

	Year	CO2	CH4	N2O	CFC12	CFC11	15-minor	Total	CO2-eq (ppm)	Total	AGGI 1990 = 1	AGGI % change
1	1979	1.027	0.406	0.104	0.092	0.039	0.031	1.699	382	0.785	Null	
2	1980	1.058	0.413	0.104	0.097	0.042	0.034	1.747	385	0.807	2.2	
3	1981	1.077	0.42	0.107	0.102	0.044	0.036	1.786	388	0.825	1.8	
4	1982	1.089	0.426	0.111	0.107	0.046	0.038	1.818	390	0.84	1.5	
5	1983	1.115	0.429	0.113	0.113	0.048	0.041	1.859	394	0.859	1.9	
6	1984	1.14	0.432	0.116	0.118	0.05	0.044	1.899	396	0.877	1.9	
7	1985	1.162	0.437	0.118	0.123	0.053	0.047	1.94	400	0.896	1.9	
8	1986	1.184	0.442	0.122	0.129	0.056	0.049	1.982	403	0.916	1.9	
9	1987	1.211	0.447	0.12	0.136	0.059	0.053	2.025	406			
10	1988	1.25	0.451	0.122	0.143	0.062	0.057	2.085	411			
11	1989	1.274	0.455	0.126	0.149	0.064	0.061	2.13	414			
12	1990	1.292	0.459	0.129	0.154	0.065	0.065	2.165	417			
13	1991	1.312	0.463	0.131	0.158	0.067	0.069	2.2	419			
14	1992	1.323	0.467	0.133	0.162	0.067	0.072	2.225	421			

Results - Input Data (8) - Output

11 of 11 Fields | Cell Viewer | 39 records displayed

Record	Year	CO2	CH4	N2O	CFC12	CFC11	15-minor	Total	CO2-eq (ppm)	Total	AGGI 1990 = 1	AGGI % change
1	1979	1.027	0.406	0.104	0.092	0.039	0.031	1.699	382	0.785	Null	
2	1980	1.058	0.413	0.104	0.097	0.042	0.034	1.747	385	0.807	2.2	
3	1981	1.077	0.42	0.107	0.102	0.044	0.036	1.786	388	0.825	1.8	
4	1982	1.089	0.426	0.111	0.107	0.046	0.038	1.818	390	0.84	1.5	
5	1983	1.115	0.429	0.113	0.113	0.048	0.041	1.859	394	0.859	1.9	
6	1984	1.14	0.432	0.116	0.118	0.05	0.044	1.899	396	0.877	1.9	
7	1985	1.162	0.437	0.118	0.123	0.053	0.047	1.94	400	0.896	1.9	
8	1986	1.184	0.442	0.122	0.129	0.056	0.049	1.982	403	0.916	1.9	

รูปที่ 4.7 การนำ Input Data Tool มาเชื่อมต่อข้อมูล

8. นำ Select Tool มาวางต่อจาก Input Data Tool จากนั้นเลือกฟิลด์ข้อมูลที่ต้องการใช้ คือ Year, CO2, CH4, N2O, CFC12, CFC11 และ 15-minor ดังรูปที่ 4.8

Options

Field	Type	Size	Ren
<input checked="" type="checkbox"/> Year	V_String	254	
<input checked="" type="checkbox"/> CO2	V_String	254	
<input checked="" type="checkbox"/> CH4	V_String	254	
<input checked="" type="checkbox"/> N2O	V_String	254	
<input checked="" type="checkbox"/> CFC12	V_String	254	
<input checked="" type="checkbox"/> CFC11	V_String	254	
<input checked="" type="checkbox"/> 15-minor	V_String	254	
<input type="checkbox"/> Total	V_String	254	
<input type="checkbox"/> CO2-eq (p...	V_String	254	
<input type="checkbox"/> AGGI 1990...	V_String	254	
<input type="checkbox"/> AGGI % ch...	V_String	254	
<input checked="" type="checkbox"/> *Unknown	Unknown	0	

TIP: To reorder multiple records, click and drag the record number.

Results - Select (9) - Output

7 of 7 Fields | Cell Viewer | 39 records displayed

Record	Year	CO2	CH4	N2O	CFC12	CFC11	15-minor
1	1979	1.027	0.406	0.104	0.092	0.039	0.031
2	1980	1.058	0.413	0.104	0.097	0.042	0.034
3	1981	1.077	0.42	0.107	0.102	0.044	0.036
4	1982	1.089	0.426	0.111	0.107	0.046	0.038
5	1983	1.115	0.429	0.113	0.113	0.048	0.041
6	1984	1.14	0.432	0.116	0.118	0.05	0.044
7	1985	1.162	0.437	0.118	0.123	0.053	0.047
8	1986	1.184	0.442	0.122	0.129	0.056	0.049
9	1987	1.211	0.447	0.12	0.136	0.059	0.053
10	1988	1.25	0.451	0.122	0.143	0.062	0.057
11	1989	1.274	0.455	0.126	0.149	0.064	0.061
12	1990	1.292	0.459	0.129	0.154	0.065	0.065
13	1991	1.312	0.463	0.131	0.158	0.067	0.069
14	1992	1.323	0.467	0.133	0.162	0.067	0.072

รูปที่ 4.8 การนำ Select Tool มาเลือกฟิลด์ที่ต้องการ

9. จากรูปที่ 4.9 เนื่องจากข้อมูลที่นำมาใช้นั้นมีข้อมูลบางส่วนที่เป็นค่า Null จึงต้องมีการกรองค่า Null ออก โดยใช้ Filter Tool และทำการตั้งค่า Configuration เพื่อไม่ให้ข้อมูลมีค่า Null โดยเลือก Customer filter จากนั้นใส่คำสั่งเพื่อกรองข้อมูลที่มี Null ดังนี้

!IsNull([Year]) AND
!IsNull([CO2]) AND
!IsNull([CH4]) AND
!IsNull([N2O]) AND
!IsNull([CFC12]) AND
!IsNull([CFC11]) AND
!IsNull([15-minor])

The screenshot shows the Alteryx workflow. On the left, the 'Select Basic or Custom Filter' dialog is open, with 'Custom filter' selected. The filter expression is: `!IsNull([Year]) AND !IsNull([CO2]) AND !IsNull([CH4]) AND !IsNull([N2O]) AND !IsNull([CFC12]) AND !IsNull([CFC11]) AND !IsNull([15-minor])`. The main workflow area shows two input tools: 'berkeley_earth_GlobalTemperatures.csv' and 'noaa_aggr_forcing.csv'. Both are connected to a 'Filter' tool. The 'Filter' tool is configured with the same custom filter expression. Below the workflow, the 'Results - Filter (10) - Out - True' table is displayed, showing 37 records.

Record	Year	CO2	CH4	N2O	CFC12	CFC11	15-minor
1	1979	1.027	0.406	0.104	0.092	0.039	0.031
2	1980	1.058	0.413	0.104	0.097	0.042	0.034
3	1981	1.077	0.42	0.107	0.102	0.044	0.036
4	1982	1.089	0.426	0.111	0.107	0.046	0.038
5	1983	1.115	0.429	0.113	0.113	0.048	0.041
6	1984	1.14	0.432	0.116	0.118	0.05	0.044
7	1985	1.162	0.437	0.118	0.123	0.053	0.047
8	1986	1.184	0.442	0.122	0.129	0.056	0.049
9	1987	1.211	0.447	0.12	0.136	0.059	0.053
10	1988	1.25	0.451	0.122	0.143	0.062	0.057
11	1989	1.274	0.455	0.126	0.149	0.064	0.061

รูปที่ 4.9 การใช้ Filter Tool กรองค่า Null

10. จากนั้นนำ Join Tool มาวางใน Workflow แล้วเชื่อมข้อมูลตารางจากสองแหล่งเข้าด้วยกันโดยใช้ Join by Specific Fields ทำการเลือก Left เป็น Year และ Right เป็น Year ดังรูปที่ 4.10

The screenshot shows the Alteryx workflow. On the left, the 'Join by Specific Fields' dialog is open. The 'Left' field is 'Year' and the 'Right' field is 'Year'. The main workflow area shows the same two input tools as before, connected to a 'Filter' tool. The 'Filter' tool is configured with the same custom filter expression. Below the workflow, the 'Results - Filter (10) - Out - True' table is displayed, showing 37 records.

Record	Year	CO2	CH4	N2O	CFC12	CFC11	15-minor
1	1979	1.027	0.406	0.104	0.092	0.039	0.031
2	1980	1.058	0.413	0.104	0.097	0.042	0.034
3	1981	1.077	0.42	0.107	0.102	0.044	0.036
4	1982	1.089	0.426	0.111	0.107	0.046	0.038
5	1983	1.115	0.429	0.113	0.113	0.048	0.041
6	1984	1.14	0.432	0.116	0.118	0.05	0.044
7	1985	1.162	0.437	0.118	0.123	0.053	0.047
8	1986	1.184	0.442	0.122	0.129	0.056	0.049
9	1987	1.211	0.447	0.12	0.136	0.059	0.053
10	1988	1.25	0.451	0.122	0.143	0.062	0.057
11	1989	1.274	0.455	0.126	0.149	0.064	0.061

รูปที่ 4.10 การตั้งค่า configuration ใน Join Tool เพื่อเชื่อมตาราง

11.เมื่อดูจากรูปที่ 4.11 จะมีฟิลด์ Year ที่ซ้ำกันเมื่อทำการ Join ข้อมูลดังนั้นจึงต้องนำฟิลด์ที่ซ้ำกันออกโดยการคลิกที่ Options และเลือก Deselect Duplicate ดังรูปที่ 4.11

	Input	Field	Type	Size	Rename	De
<input checked="" type="checkbox"/>	Left	Year	V_String	254		Text
<input checked="" type="checkbox"/>	Left	Avg_Land...	Double	8		
<input type="checkbox"/>	Right	Year	V_String	254	Right_Year	
<input checked="" type="checkbox"/>	Right	CO2	V_String	254		
<input checked="" type="checkbox"/>	Right	CH4	V_String	254		
<input checked="" type="checkbox"/>	Right	N2O	V_String	254		
<input checked="" type="checkbox"/>	Right	CFC12	V_String	254		
<input checked="" type="checkbox"/>	Right	CFC11	V_String	254		
<input checked="" type="checkbox"/>	Right	15-minor	V_String	254		
<input checked="" type="checkbox"/>		*Unknown	Unknown	0		Dyn

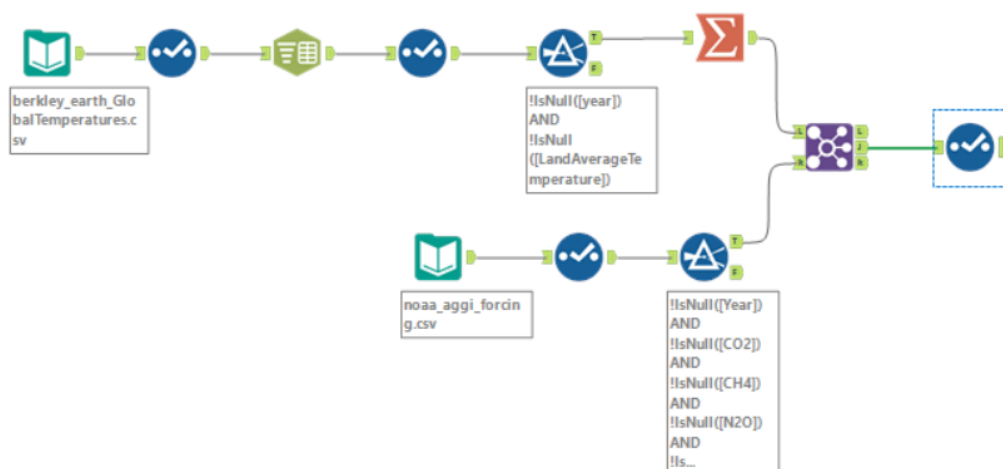
รูปที่ 4.11 เป็นส่วนของ configuration ด้านล่างใน Join Tool

ผลลัพธ์หลังการเชื่อมข้อมูลจากทั้ง 2 ตารางจะได้ทั้งหมด 9 Fields 37 records ดังรูปที่ 4.12

Record	Year	Avg_LandAverageTemperature	Right_Year	CO2	CH4	N2O	CFC12	CFC11	15-minor
1	1979	8.25	1979	1.027	0.406	0.104	0.092	0.039	0.031
2	1980	8.416667	1980	1.058	0.413	0.104	0.097	0.042	0.034
3	1981	8.666667	1981	1.077	0.42	0.107	0.102	0.044	0.036
4	1982	8.166667	1982	1.089	0.426	0.111	0.107	0.046	0.038
5	1983	8.416667	1983	1.115	0.429	0.113	0.113	0.048	0.041
6	1984	8.166667	1984	1.14	0.432	0.116	0.118	0.05	0.044
7	1985	8.083333	1985	1.162	0.437	0.118	0.123	0.053	0.047
8	1986	8.166667	1986	1.184	0.442	0.122	0.129	0.056	0.049
9	1987	8.5	1987	1.211	0.447	0.12	0.136	0.059	0.053
10	1988	8.583333	1988	1.25	0.451	0.122	0.143	0.062	0.057

รูปที่ 4.12 ผลลัพธ์จากการเชื่อมตาราง

12.ใช้ Select Tool ดังรูปที่ 4.13 ในการเลือกฟิลด์ที่ต้องการ คือ Year, CO2, CH4, N2O, CFC12, CFC11 และ 15-minor แล้วเปลี่ยน Type จาก String เป็น Float ยกเว้นฟิลด์ Year ดังที่แสดงในรูปที่ 4.14



รูปที่ 4.13 การนำ Select Tool มาใช้ในการเลือกฟิลด์

Options ▾ | ↑ ↓ TIP: To reorder multiple rows: select, right-click and drag

	Field	Type	Size	Rename	Description
<input checked="" type="checkbox"/>	Year	V_String	254		Text ToColumn...
<input checked="" type="checkbox"/>	Avg_Land...	Float	4		
<input checked="" type="checkbox"/>	CO2	Float	4		
<input checked="" type="checkbox"/>	CH4	Float	4		
<input checked="" type="checkbox"/>	N2O	Float	4		
<input checked="" type="checkbox"/>	CFC12	Float	4		
<input checked="" type="checkbox"/>	CFC11	Float	4		
<input checked="" type="checkbox"/>	15-minor	Float	4		
<input type="checkbox"/>	Right_Year...	Unknown	N/A		
<input type="checkbox"/>	Total (Missi...	Unknown	N/A		
<input type="checkbox"/>	CO2-eq (p...	Unknown	N/A		
<input type="checkbox"/>	AGGI 1990...	Unknown	N/A		
<input type="checkbox"/>	AGGI % ch...	Unknown	N/A		
<input checked="" type="checkbox"/>	*Unknown	Unknown	0		Dynamic or U...

รูปที่ 4.14 การ configuration ค่าใน Select Tool

ผลลัพธ์ที่ได้หลังจากการใช้ Select Tool จะได้ข้อมูลทั้งหมด 8 Fields 37 records ดังรูปที่ 4.15

8 of 8 Fields ▾ ✓ | Cell Viewer ▾ 37 records displayed | ↑ ↓ | Search

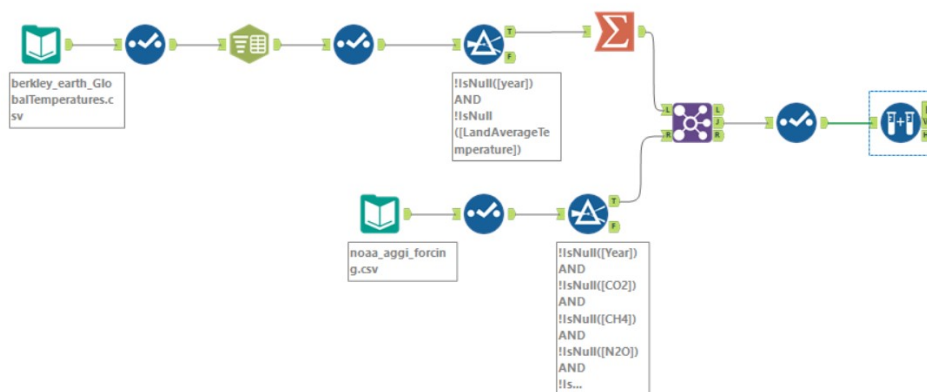
Record	Year	Avg_LandAverageTemperature	CO2	CH4	N2O	CFC12	CFC11	15-minor
1	1979	8.25	1.027	0.406	0.104	0.092	0.039	0.031
2	1980	8.416667	1.058	0.413	0.104	0.097	0.042	0.034
3	1981	8.666667	1.077	0.42	0.107	0.102	0.044	0.036
4	1982	8.166667	1.089	0.426	0.111	0.107	0.046	0.038
5	1983	8.416667	1.115	0.429	0.113	0.113	0.048	0.041
6	1984	8.166667	1.14	0.432	0.116	0.118	0.05	0.044
7	1985	8.083333	1.162	0.437	0.118	0.123	0.053	0.047
8	1986	8.166667	1.184	0.442	0.122	0.129	0.056	0.049
9	1987	8.5	1.211	0.447	0.12	0.136	0.059	0.053
10	1988	8.583333	1.25	0.451	0.122	0.143	0.062	0.057
11	1989	8.416667	1.274	0.455	0.126	0.149	0.064	0.061
12	1990	8.666667	1.292	0.459	0.129	0.154	0.065	0.065

รูปที่ 4.15 ผลลัพธ์จากการใช้ Select Tool

5.วิธีการแก้ไขปัญหาด้วย Data Mining

หลังจากได้ทำการจัดเตรียมและทำความสะอาดข้อมูลแล้ว จึงจะสามารถนำข้อมูลมาทำการพยากรณ์ (Prediction) ด้วยเทคนิค Linear Regression ดังขั้นตอนต่อไปนี้

1.นำ Create Samples Tool ดังรูปที่ 5.1 มาวางต่อจาก Select Tool เพื่อใช้แบ่งข้อมูลออกเป็น 2 ชุด



รูปที่ 5.1 การนำ Create Sample Tool มาวางใน workflow

2.กำหนดค่า configuration ให้ Estimation sample percent เป็น 80 และ Validation sample percent เป็น 20 เพื่อแบ่งชุดข้อมูลออกเป็น 2 ชุด คือ Train และ Test ดังรูปที่ 5.2

รูปที่ 5.2 การ configuration ใน Create Samples Tool

จากรูปที่ 5.3 จะแสดงผลลัพธ์ของชุดข้อมูลสำหรับใช้ Train จะมีทั้งหมด 8 Fields 30 records

8 of 8 Fields		Cell Viewer		30 records displayed		Search		
Record	Year	Avg_LandAverageTemperature	CO2	CH4	N2O	CFC12	CFC11	15-minor
1	1979	8.25	1.027	0.406	0.104	0.092	0.039	0.031
2	1980	8.416667	1.058	0.413	0.104	0.097	0.042	0.034
3	1981	8.666667	1.077	0.42	0.107	0.102	0.044	0.036
4	1982	8.166667	1.089	0.426	0.111	0.107	0.046	0.038
5	1983	8.416667	1.115	0.429	0.113	0.113	0.048	0.041
6	1984	8.166667	1.14	0.432	0.116	0.118	0.05	0.044
7	1985	8.083333	1.162	0.437	0.118	0.123	0.053	0.047
8	1986	8.166667	1.184	0.442	0.122	0.129	0.056	0.049
9	1987	8.5	1.211	0.447	0.12	0.136	0.059	0.052

รูปที่ 5.3 ผลลัพธ์จากชุดข้อมูล Train

จากรูปที่ 5.4 จะแสดงผลชุดข้อมูลสำหรับใช้ Test จะมีทั้งหมด 8 Fields 7 records

8 of 8 Fields ▾ ✓		Cell Viewer ▾ 7 records displayed		↑ ↓		Search		
Record	Year	Avg_LandAverageTemperature	CO2	CH4	N2O	CFC12	CFC11	15-minor
1	1989	8.416667	1.274	0.455	0.126	0.149	0.064	0.061
2	1990	8.666667	1.292	0.459	0.129	0.154	0.065	0.065
3	1993	8.25	1.334	0.467	0.133	0.164	0.068	0.074
4	1997	8.75	1.426	0.474	0.142	0.171	0.067	0.079
5	2003	9.083333	1.6	0.483	0.157	0.174	0.064	0.088
6	2009	9.25	1.76	0.489	0.172	0.171	0.061	0.103
7	2013	9.25	1.882	0.496	0.184	0.167	0.059	0.114

รูปที่ 5.3 ผลลัพธ์จากชุดข้อมูล Test

3. นำ Linear Regression Tool มาวางต่อจาก Create samples Tool ในขา E เพื่อทำการวิเคราะห์โมเดล ดังที่แสดงในรูปที่ 5.4



รูปที่ 5.4 การนำ Linear Regression Tool มาวางใน workflow

ทำการตั้งค่า Configuration โดยการตั้งชื่อโมเดลว่า Linear_Regression จากนั้นทำการ Select the target variable (เลือกค่าเป้าหมาย) คือ Avg_landAverageTemperature แล้ว Select the predictor variables (เลือกตัวแปรต้น) ที่คิดว่าจะส่งผลต่อตัวแปรตามหรือ ค่าเป้าหมายที่ได้เลือกไว้ ซึ่งทางผู้จัดทำได้เลือกตัวแปรต้น คือ CO2, CH4, N2O, CFC12, CFC11 และ 15-minor ดังรูปที่ 5.5

Setup

Model name

Linear_Regression

Select the target variable

Avg_LandAverageTemperature ▾

Select the predictor variables

Selected: 6 Fields: 8 Show: All Selected

<input type="checkbox"/>	
<input type="checkbox"/>	Year
<input type="checkbox"/>	Avg_LandAverageTemperature
<input checked="" type="checkbox"/>	CO2
<input checked="" type="checkbox"/>	CH4
<input checked="" type="checkbox"/>	N2O
<input checked="" type="checkbox"/>	CFC12
<input checked="" type="checkbox"/>	CFC11
<input checked="" type="checkbox"/>	15-minor

รูปที่ 5.5 การตั้งค่าใน Linear Regression Tool

4.จากรูปที่ 5.6 ได้นำ Browse Tool มาต่อกับขา R ของ Linear Regression Tool เพื่อดูผลลัพธ์ที่เป็นรายงานทางสถิติต่างๆ ในรูปที่ 5.7 และ 5.8



รูปที่ 5.6 การนำ Browse Tool มาวางใน workflow

Report

Report for Linear Model Linear_Regression

Basic Summary

Call:

```
lm(formula = Avg_LandAverageTemperature ~ CO2 + CH4 + N2O + CFC12 + CFC11 + X15.minor, data = the.data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.2647	-0.1089	-0.0116	0.1208	0.2604

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.5415	4.984	1.51328	0.14383
CO2	7.8459	2.636	2.97673	0.00675 **
CH4	-0.8047	15.772	-0.05102	0.95975
N2O	-68.1715	27.180	-2.50815	0.01964 *
CFC12	11.4460	9.318	1.22838	0.23173
CFC11	-17.6024	30.082	-0.58516	0.56414
X15.minor	-9.6098	17.261	-0.55674	0.58308

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.16758 on 23 degrees of freedom

Multiple R-squared: 0.8347, Adjusted R-Squared: 0.7916

F-statistic: 19.36 on 6 and 23 degrees of freedom (DF), p-value 6.212e-08

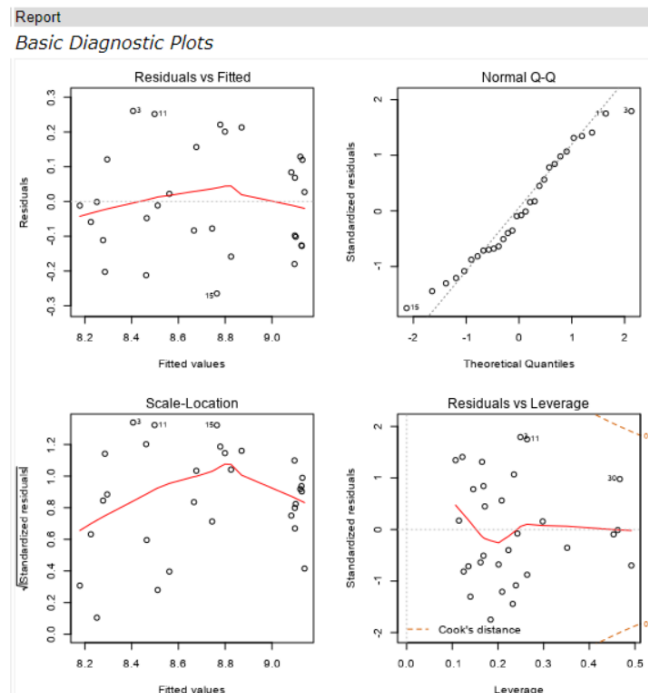
Type II ANOVA Analysis

Response: Avg_LandAverageTemperature

	Sum Sq	DF	F value	Pr(>F)
CO2	0.25	1	8.86	0.00675 **
CH4	0	1	0	0.95975
N2O	0.18	1	6.29	0.01964 *
CFC12	0.04	1	1.51	0.23173
CFC11	0.01	1	0.34	0.56414
X15.minor	0.01	1	0.31	0.58308
Residuals	0.65	23		

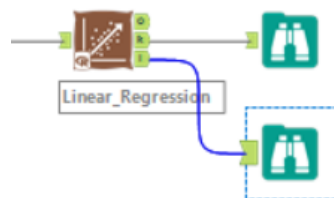
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

รูปที่ 5.7 รายงานทางสถิติโดยสรุปต่างๆของสมการทำนาย



รูปที่ 5.7 รายงานของกราฟ Basic Diagnostic Plots

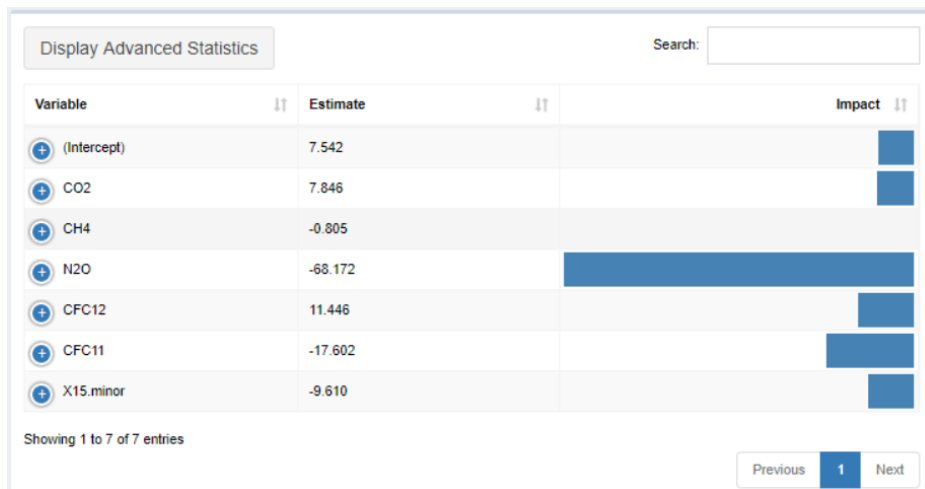
5. นำ Browse Tool มาต่อกับขา I ของ Linear Regression Tool ดังรูปที่ 5.8 เพื่อดูผลลัพธ์ซึ่งผลลัพธ์ของขา I นั้นจะคล้ายกับขา R แต่จะแยกแต่ละเรื่องของค่าทางสถิติดังนี้ Summary(ดังรูปที่ 5.9 และ 5.10), Model Performance(ดังรูปที่ 5.11 และ 5.12), Diagnostics(ดังรูปที่ 5.11)



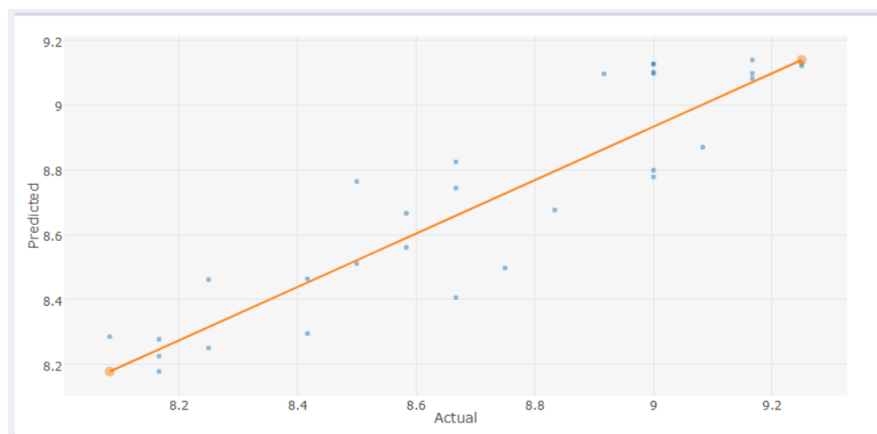
รูปที่ 5.8 การนำ Browse Tool มาวางใน workflow

✓ R SQUARED 0.835	✓ ADJUSTED R SQUARED 0.792
✓ MEAN ABSOLUTE ERROR 0.125	✓ MEAN ABSOLUTE PERCENT ERR... 0.014
✓ MEAN SQUARED ERROR 0.022	✓ ROOT MEAN SQUARED ERROR 0.147
✓ F-STATISTIC 19.36 on 6 and 23 degrees of freedom	✓ RESIDUAL STANDARD ERROR 0.168 on 24 degrees of freedom

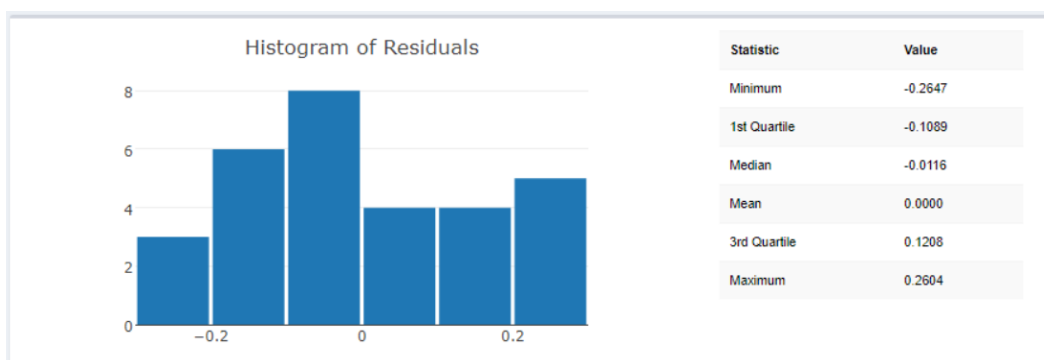
รูปที่ 5.9 รายงานแสดงค่าสถิติพรรณนาสรุปต่างๆ



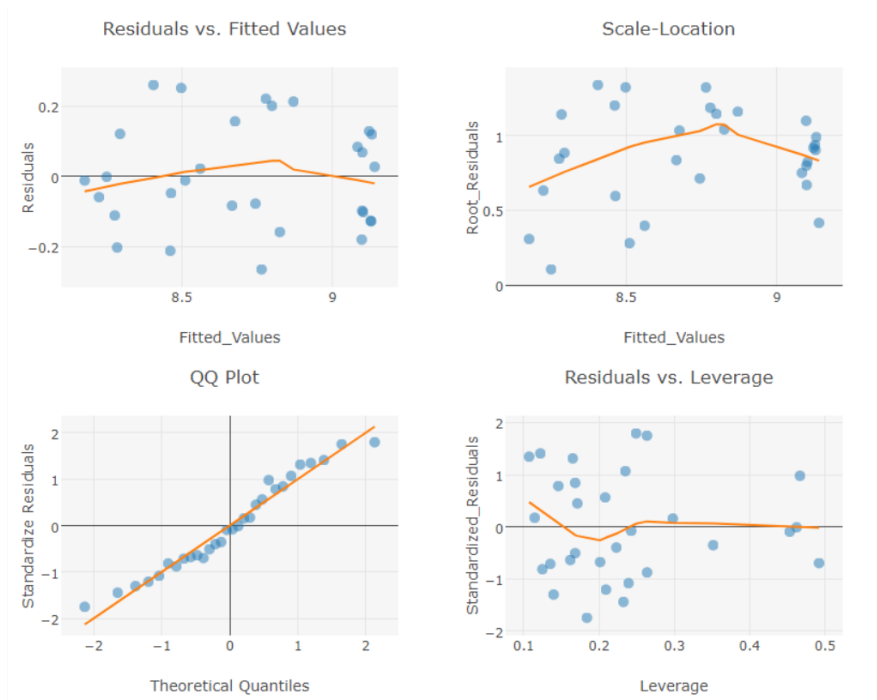
รูปที่ 5.10 รายงานแสดงค่า coefficient ของสมการทำนาย



รูปที่ 5.11 Scatter Plot แสดงความสัมพันธ์ระหว่างค่า Predict กับค่า Actual

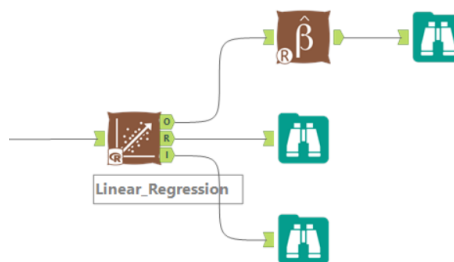


รูปที่ 5.12 Histogram ของค่าความคลาดเคลื่อนของสมการทำนาย



รูปที่ 5.13 กราฟรายงานของสถิติแบบ Diagnostics

6. หลังจากนั้นจะต้องตรวจสอบค่า Coefficients ที่นำมาสร้างสมการการทำนายโดยจะต้องทำการรันไฟล์ที่ชื่อว่า Model+Coefficients.yxzp ซึ่งเป็น Alteryx Macro หลังจากนั้นทำการนำ Model Coefficients Tool มาเชื่อมต่อกับขา O ของ Linear Regression Tool และนำ Browse มาเชื่อมต่อกับ Model Coefficients Tool ดังรูปที่ 5.14 เพื่อดูผลลัพธ์หลังจากใช้ Browse Tool ดังรูปที่ 5.15



รูปที่ 5.14 การนำ Model Coefficients Tool และ Browse Tool มาวางใน workflow

2 of 2 Fields ▾ ✓		Cell Viewer ▾ 7 records	
Record	Variable	Coefficient	
1	(Intercept)	7.541508	
2	CO2	7.845904	
3	CH4	-0.804697	
4	N2O	-68.17152	
5	CFC12	11.446014	
6	CFC11	-17.602419	
7	X15.minor	-9.60975	

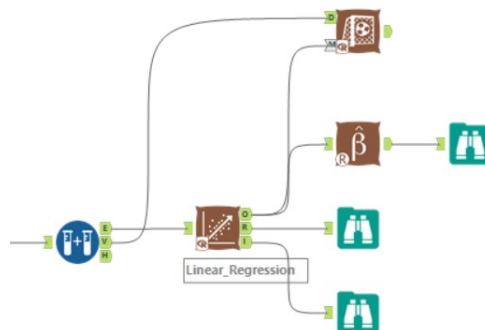
รูปที่ 5.15 ผลลัพธ์ Model Coefficients Tool เมื่อเชื่อมกับ Browse Tool

จากรูปที่ 5.10 และ 5.15 จะเห็นค่า Coefficient ของแต่ละตัวแปร ดังนั้นสมการทำนายในการคำนวณจากโมเดลนี้ คือ

$$Y = 7.541508 + (7.845904)X_1 + (-0.804697)X_2 + (-68.17152)X_3 + (11.4460114)X_4 + (-17.602419)X_5 + (-9.60975)X_6$$

$$\text{หรือ Avg_landAverageTemperature} = 7.541508 + (7.845904)\text{CO2} + (-0.804697)\text{CH4} + (-68.17152)\text{N2O} + (11.4460114)\text{CFC12} + (-17.602419)\text{CFC11} + (-9.60975)\text{X15.minor}$$

7. นำ Score Tool มาวางเชื่อมต่อ โดยนำ Input Data มาจากข้อมูล 20% ที่แบ่งไว้จาก Create Samples Tool ในขา V ต่อกับขา D และ ขา O จาก Linear Regression Tool ให้ต่อที่ขา M ดังรูปที่ 5.16

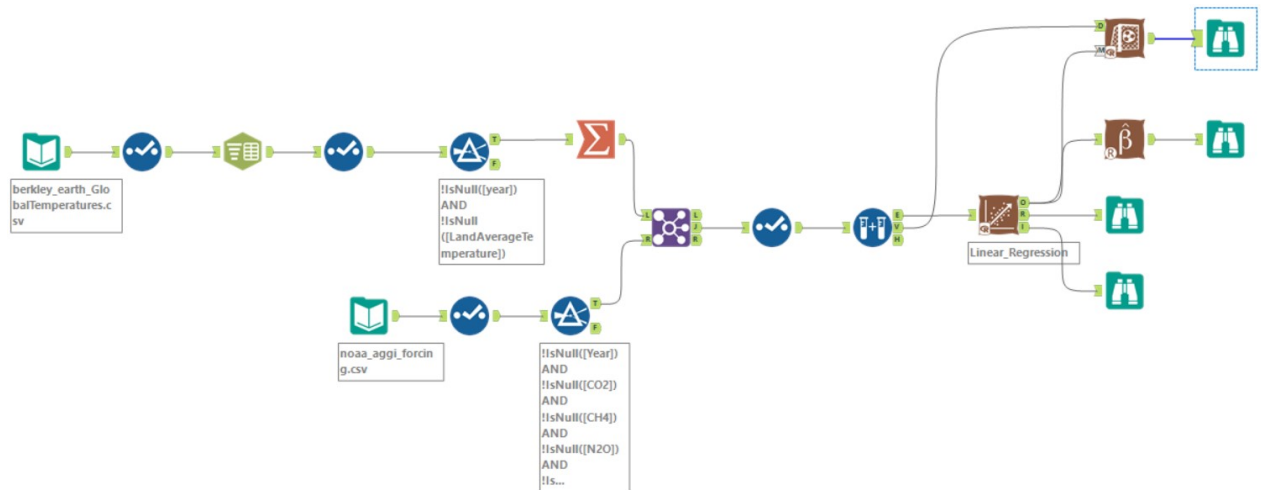


รูปที่ 5.16 การนำ Score Tool มาวางใน workflow

ทำการ configuration ใน Score Tool โดยตั้งชื่อฟิลด์ใหม่ว่า Predict ใน Configure Local Model Options ดังที่แสดงในรูปที่ 5.17

รูปที่ 5.17 การตั้งค่าใน Score Tool

8. นำ Browse Tool มาเชื่อมต่อจาก Score Tool เพื่อดูผลลัพธ์หลังจากนั้นจะได้ workflow สุดท้าย ดังรูปที่ 5.18



รูปที่ 5.18 การนำ Browse Tool มาวางใน workflow

ผลลัพธ์หลังจากใช้ Browse Tool มาเชื่อมต่อจาก Score Tool จะได้ข้อมูลทั้งหมด 9 Fields 7 records ดังที่แสดงในรูปที่ 5.19

9 of 9 Fields Cell Viewer 7 records displayed, 2147 bytes Search									
Record	Year	Avg_LandAverageTemperature	CO2	CH4	N2O	CFC12	CFC11	X15.minor	Predict
1	1989	8.416667	1.274	0.455	0.126	0.149	0.064	0.061	8.574148
2	1990	8.666667	1.292	0.459	0.129	0.154	0.065	0.065	8.50883
3	1993	8.25	1.334	0.467	0.133	0.164	0.068	0.074	8.534398
4	1997	8.75	1.426	0.474	0.142	0.171	0.067	0.079	8.68672
5	2003	9.083333	1.6	0.483	0.157	0.174	0.064	0.088	9.02275
6	2009	9.25	1.76	0.489	0.172	0.171	0.061	0.103	9.125017
7	2013	9.25	1.882	0.496	0.184	0.167	0.059	0.114	9.14224

รูปที่ 5.19 ผลลัพธ์จากการใช้ Browse Tool มาเชื่อมกับ Score Tool

6.บทสรุป

ในการสร้างการทำนายด้วยเทคนิคการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression) จะเห็นได้ผลลัพธ์ของค่า coefficient จากรูปที่ 5.10 ออกมาเป็นสมการทำนายดังนี้

$$Y = 7.541508 + (7.845904)X_1 + (-0.804697)X_2 + (-68.17152)X_3 + (11.4460114)X_4 + (-17.602419)X_5 + (-9.60975)X_6$$

$$\text{หรือ Avg_landAverageTemperature} = 7.541508 + (7.845904)CO_2 + (-0.804697)CH_4 + (-68.17152)N_2O + (11.4460114)CFC_{12} + (-17.602419)CFC_{11} + (-9.60975)X_{15.minor}$$

และจากรูปที่ 5.19 จะเห็นผลลัพธ์ของค่า Predict ที่ได้นั้นมีค่าค่อนข้างใกล้เคียงกับค่า Avg_landAverageTemperature แสดงให้เห็นว่าโมเดลมีความแม่นยำพอสมควร โดยสามารถดูเปอร์เซ็นต์ความแม่นยำของการทำนายนี้ได้จากรูปที่ 5.9 ด้วยค่า Adjusted R Squared ที่มีค่าเท่ากับ 0.792 ซึ่งหมายความว่า สมการการทำนายนี้สามารถทำนายค่า Predict ด้วยความแม่นยำสูงสุด 79.2% ซึ่งถือว่าอยู่ในระดับที่ดีจึงทำให้สามารถนำสมการการทำนายนี้ไปใช้พยากรณ์อุณหภูมิเฉลี่ยของโลกในอนาคตได้ เพื่อนำไปช่วยวิเคราะห์ในการลดปริมาณก๊าซเรือนกระจกในปัจจุบัน และช่วยในการวางแผนการจัดการทรัพยากรธรรมชาติในอนาคตอย่างเหมาะสม

บรรณานุกรม

Austin Schwinn. 2017. **Global Temp & Greenhouse Gas**. [Online].

Available : <https://data.world/amschwinn/global-temp-greenhouse-gas>

