



Rainfall prediction

โดย

62010282 นายณัฐภัทร อรุณกิจเจริญ

นำเสนอ

ดร.วัชระ ฉัตรวิริยะ

รายงานนี้เป็นส่วนหนึ่งของรายวิชา 01076585 Data Mining

คณะวิศวกรรมศาสตร์ สาขาวิชาวิศวกรรมคอมพิวเตอร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาคเรียนที่ 2 ปีการศึกษา 2

1. การเตรียมข้อมูล

1.1 จาก Dataset ที่ได้รับจากอาจารย์ ประกอบด้วย

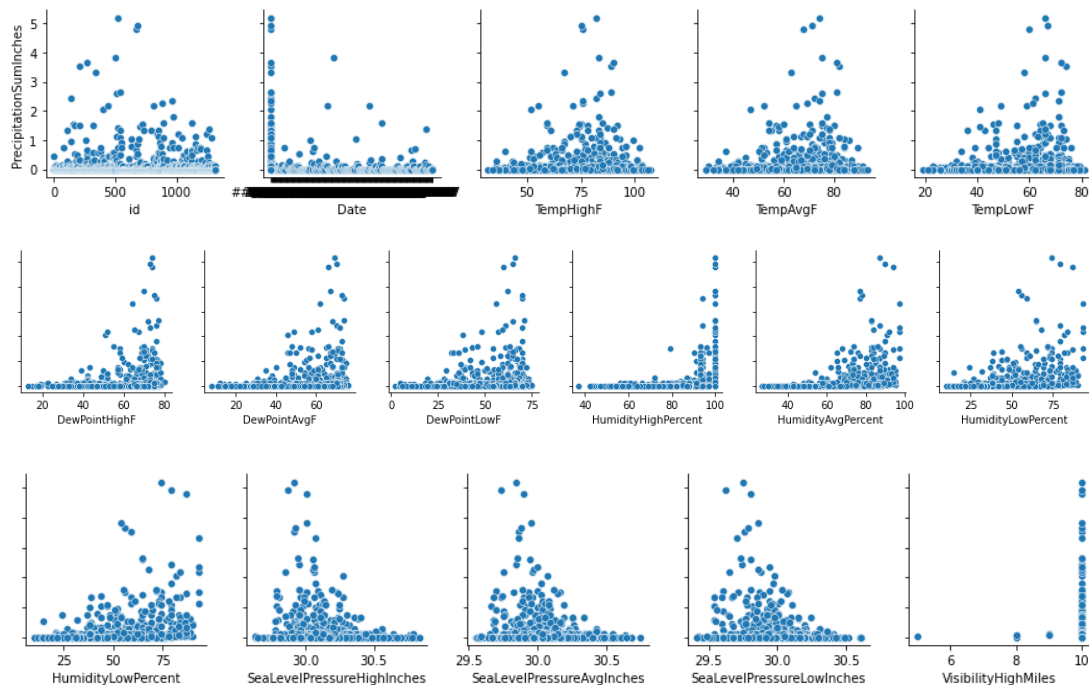
- a.) Date คือ วันที่
- b.) Temperature คือ อุณหภูมิภาค โดยประกอบด้วย
- c.) Dew point temperature คือ อุณหภูมิจุดน้ำค้าง
- d.) Humidity คือ ค่าความชื้น
- e.) Sea Level Pressure คือ ค่าความกดอากาศ
- f.) Visibility คือ ระดับทัศนวิสัย
- g.) Wind คือ ระดับความแรงลม
- h.) PrecipitationSumInches คือ ปริมาณน้ำฝน

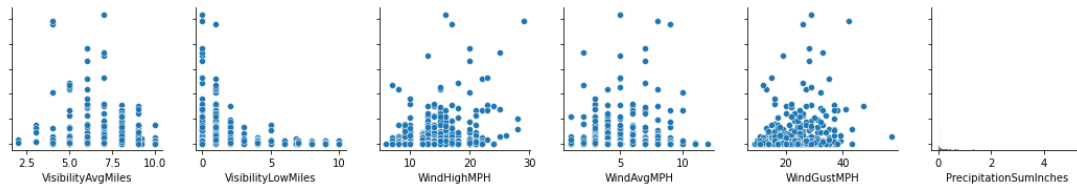
1.2 การจัดการข้อมูลที่ขาดหาย

ใช้วิธีการเปลี่ยนข้อมูลที่ขาดหายเป็นค่าเฉลี่ยตาม column ของข้อมูลนั้นๆ

2. การปรับจูนพารามิเตอร์

จากข้อมูลหาตาราง Correlation Coefficient ได้ดังนี้





โดยหลังจากการหาพบว่า ได้ค่าความสัมพันธ์ดังนี้

	PrecipitationSumInches
VisibilityAvgMiles	-0.428700
VisibilityLowMiles	-0.428445
SeaLevelPressureLowInches	-0.134684
SeaLevelPressureAvgInches	-0.123756
SeaLevelPressureHighInches	-0.106216
TempHighF	-0.068840
TempAvgF	-0.017103
id	-0.010582
VisibilityHighMiles	0.011289
TempLowF	0.039878
WindAvgMPH	0.042782
DewPointLowF	0.130261
DewPointAvgF	0.132137
DewPointHighF	0.142229
WindGustMPH	0.205632
HumidityHighPercent	0.229230
WindHighMPH	0.247057
HumidityAvgPercent	0.354454
HumidityLowPercent	0.369559

จากตารางพบว่า ค่าที่มีความสัมพันธ์กับข้อมูลมีดังนี้

- VisibilityAvgMiles
- VisibilityLowMiles
- HumidityAvgPercent

d) HumidityLowPercent

จากนั้นสร้างโมเดล

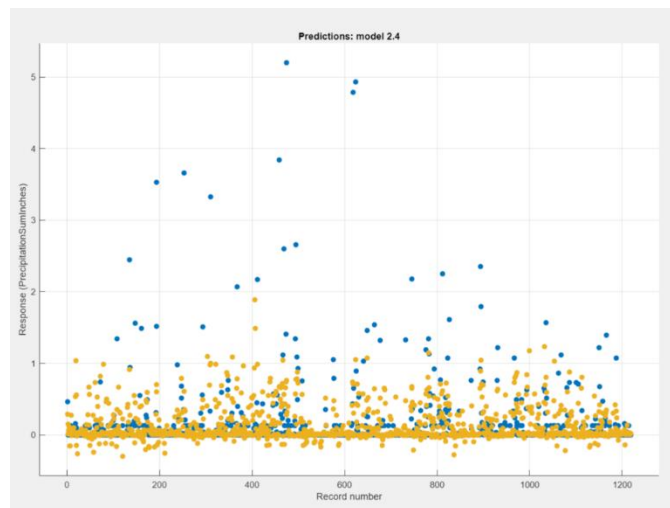
3. ประสิทธิภาพของโมเดลที่สร้าง

☆ 2.1 Linear Regression	RMSE (Validation): 0.38779
Last change: Linear	4/4 features
☆ 2.2 Linear Regression	RMSE (Validation): 0.3751
Last change: Interactions Linear	4/4 features
☆ 2.3 Linear Regression	RMSE (Validation): 0.44447
Last change: Robust Linear	4/4 features
☆ 2.4 Stepwise Linear Regression	RMSE (Validation): 0.37405
Last change: Stepwise Linear	4/4 features

ดังนั้นจึงเลือกเทคนิค Stepwise Linear

4. เหตุผลที่เลือกโมเดล

เนื่องจาก Model ที่ใช้ Stepwise Linear มีค่า RMSE น้อยที่สุด



5. ผลการทำนายปริมาณน้ำฝนของข้อมูลทดสอบ

id	Predict
12	0.0623
32	0.0253
34	0.2469
46	0.1761
58	0.7799
63	0.0253
95	0.0412

130	0.0356
153	0.0969
158	0.9622
166	-0.0169
167	-0.022
169	-0.013
182	0.0227
194	0.077
207	0.0585
211	-0.0147
229	0.0229
251	-0.0015
287	0.0177
292	-0.0092
295	0.825
299	0.0241
307	-0.0092
310	0.0397
311	-0.0015
322	0.0164
355	-0.0337
360	0.0286
363	0.8174
416	0.0241
435	0.5046
455	1.0193
459	0.0228
460	-0.0156
466	0.107
474	0.0233
476	-0.0105
492	0.4819
494	-0.0322
496	0.0359
500	0.2017
510	0.5736
511	0.6138
514	0.268
517	0.155
525	0.3805
533	-0.0092
541	0.7677
614	0.0036
616	0.0151
626	0.0074
642	0.023

666	0.0177
668	0.0177
683	0.5283
703	0.0113
705	0.2267
721	0.7225
736	0.1555
739	0.1691
755	-0.0233
793	0.0963
797	0.0164
805	0.0177
813	0.0585
814	-0.0194
818	0.5431
874	0.371
885	0.4776
907	0.0674
914	-0.0079
919	0.0035
938	0.0229
950	0.6883
975	0.2561
984	-0.0169
992	0.0623
994	-0.013
1006	0.0539
1027	-0.0105
1046	0.4253
1050	0.0727
1056	0.0636
1062	0.0113
1078	0.2369
1079	0.5577
1086	0.2824
1106	0.0151
1137	0.0305
1176	-0.0105
1190	0.2025
1191	0.0177
1194	-0.038
1205	-0.0156
1232	0.0241
1239	-0.029
1291	0.0049
1309	0.0151

1310	0.0151
------	--------