An abstract 3D composition featuring several geometric shapes: a blue cube, a brown cube, a dark blue cube, and a dark blue cylinder. Two bright yellow circles are also present. The shapes are arranged in a dynamic, overlapping manner against a dark blue background with a subtle gradient. A thin white horizontal line is positioned above the main text.

Prediction of subway usage by time according to weather

2023.08.04

San Gwon, Joeun Kim, Hyeonjeong Kim, Yejin Lee

Team Joeun Day

1 Fault of previous datasets

2 MLR result

3 Random Forest

Part 1

Fault of previous datasets

Select the continuous years

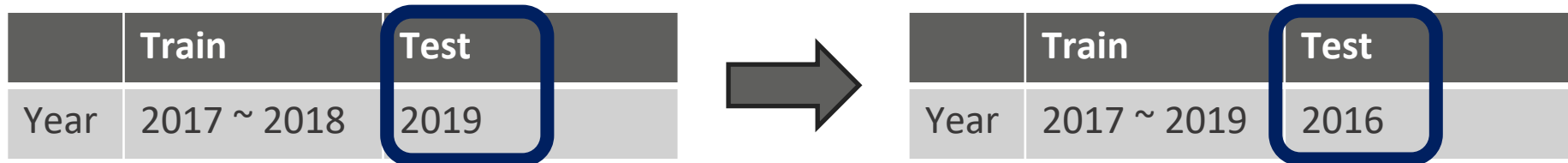
2018-12-31	-1.6	0	1.9	22	-4.20302	17
2018-12-31	-2.4	0	1.7	27	-4.83605	18
2018-12-31	-3.1	0	2.1	43	-6.2209	19
2018-12-31	-3.7			39	-3.7	20
2018-12-31	-4.6	0	1.1	44	-4.6	21
2018-12-31	-5.4	0	1.3	46	-5.4	22
2018-12-31	-5.2	0	1.6	47	-7.88648	23
2019-01-02	-8.5	0	0.8	57	-8.5	5
2019-01-02	-8.5			57	-8.5	6
2019-01-02	-8.5	0	0.7	61	-8.5	7
2019-01-02	-8.7	0	0.8	63	-8.7	8

Last Train Data

First Test Data

For this reason, higher accuracy is obtained.

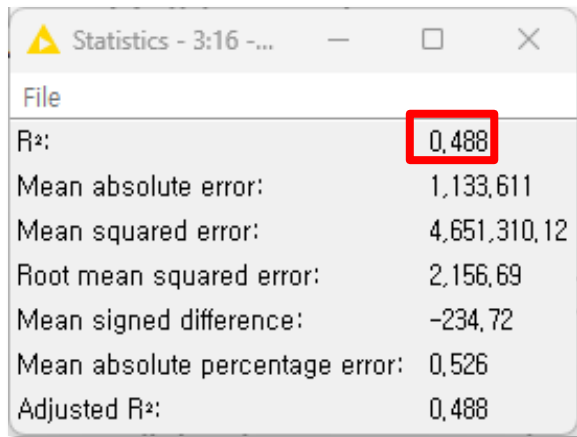
Part 1 Fault of previous datasets



Change the period of the train data and test data

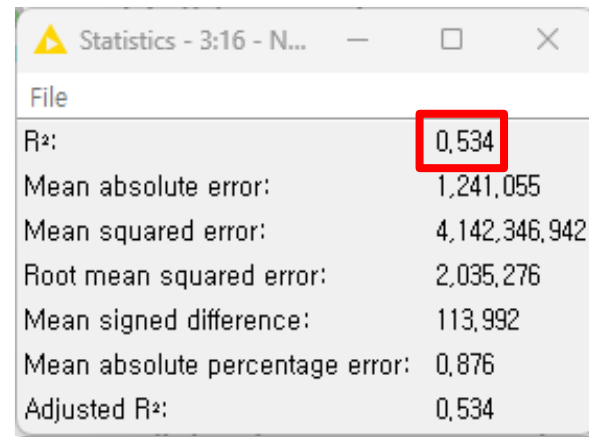
Part 1 Fault of previous datasets

summer



Statistics - 3:16 - ...	
File	
R²:	0.488
Mean absolute error:	1,133,611
Mean squared error:	4,651,310,12
Root mean squared error:	2,156,69
Mean signed difference:	-234,72
Mean absolute percentage error:	0,526
Adjusted R²:	0,488

winter



Statistics - 3:16 - N...	
File	
R²:	0.534
Mean absolute error:	1,241,055
Mean squared error:	4,142,346,942
Root mean squared error:	2,035,276
Mean signed difference:	113,992
Mean absolute percentage error:	0,876
Adjusted R²:	0,534

Weather	R square
Winter	0.2116
Fall	0.0516
Spring	0.0434
Summer	0.0411

Seasonal MLR results that do not include historical usage

Result of MLR in summer and winter

$$0.534 - 0.488 = 0.046 \quad \text{a very small number}$$

Part 1

Fault of previous datasets

Correlation between usage and independent variables

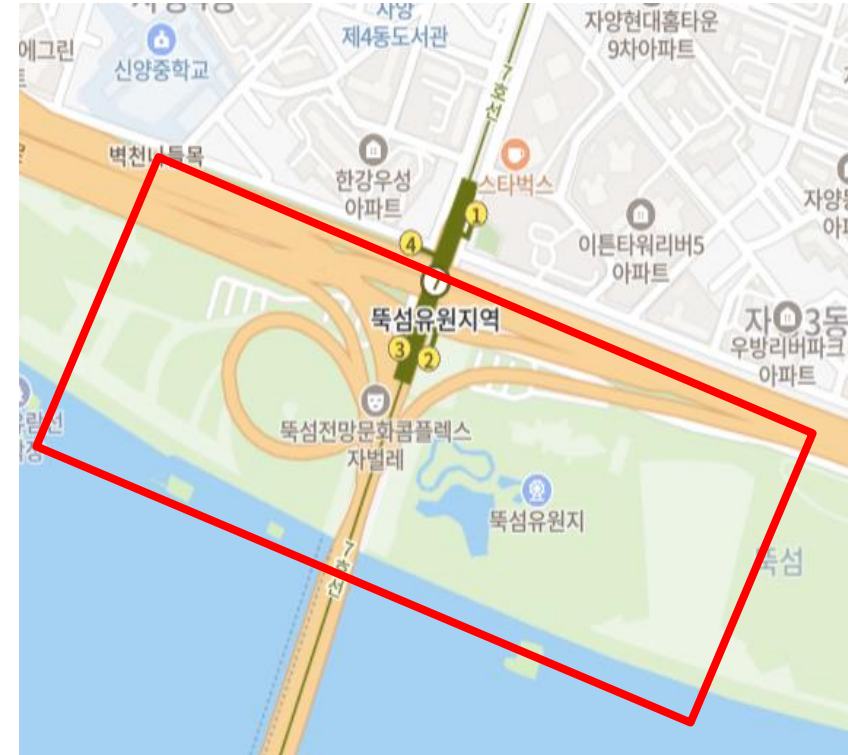
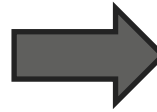
B	A Usage	F
Usage(week)		0.988
Usage(1hour)		0.669
Hour		0.586
Wind(1hour)		0.397
Wind		0.295
Wind(week)		0.287
Temp(1hour)		0.246
Temp(week)		0.152
Temp		0.152
Rain(week)		0.017
Rain		-0.005
Rain(1hour)		-0.017
Humidity(week)		-0.143
Humidity		-0.165
Humidity(1hour)		-0.260

Part 1 Fault of previous datasets

After deciding on a station with a large number of Han River users, not a previous station near the company, the pretreatment process was performed.



Euljiro Entrance Station:
Company-intensive area



Ttukseom Amusement Park Station:
Many visitors to the Han River and regular visitors

Part 2

MLR result

2016 R square value

Weather	\bar{r}
Sensible_Temp+Humidity	0.897
Sensible_Temp	0.892
normal	0.884
Sensible_Temp+Humidity+Rain	0.873
Humidity	0.869
Rain	0.816
Sensible_Temp+Rain	0.815
Rain+Humidity	0.731
Wind	0.210
Rain+Wind	-0.167
Humidity+Wind	-0.219
Rain+Humidity+Wind	-0.610
Sensible_Temp+Wind	-1.488
Sensible_Temp+Humidity+Wind	-1.667
All	-2.266
Sensible_Temp+Rain+Wind	-2.753

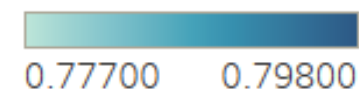
2019 R square value

Weather	\bar{r}
All	0.798
Sensible_Temp+Humidity+Rain	0.797
Sensible_Temp+Humidity+Wind	0.796
Sensible_Temp+Humidity	0.795
Rain+Humidity+Wind	0.795
Rain+Humidity	0.795
Humidity	0.793
Humidity+Wind	0.792
Sensible_Temp+Rain+Wind	0.789
Sensible_Temp+Rain	0.788
Rain+Wind	0.787
Rain	0.787
Sensible_Temp+Wind	0.78
Sensible_Temp	0.779
Wind	0.777
normal	0.777

2016 Value



2019 Value



Part 2

MLR result

2016 up R

All data included
Hour and Usage

Weather	
All	0.899
Sensible_Temp+Humidity+Rain	0.899
Sensible_Temp+Humidity+Wind	0.899
Sensible_Temp+Humidity	0.898
Sensible_Temp	0.892
Sensible_Temp+Rain	0.892
Sensible_Temp+Rain+Wind	0.892
Sensible_Temp+Wind	0.892
Humidity	0.891
Humidity+Wind	0.891
Rain+Humidity	0.891
Rain+Humidity+Wind	0.891
Rain	0.885
normal	0.884
Rain+Wind	0.884
Wind	0.884

Proceed in combination to determine
which weather data affect

Part 2

MLR result

All data included
Hour and Usage

2016 down R

Weather	F
Sensible_Temp+Humidity	0.891
Sensible_Temp+Humidity+Wind	0.891
Sensible_Temp	0.882
Humidity	0.880
Humidity+Wind	0.879
Sensible_Temp+Wind	0.879
normal	0.870
Wind	0.867
All	0.483
Sensible_Temp+Humidity+Rain	0.447
Rain+Humidity+Wind	0.139
Rain+Humidity	0.049
Sensible_Temp+Rain	-0.792
Sensible_Temp+Rain+Wind	-0.901
Rain+Wind	-1.124
Rain	-1.309

Proceed in combination to determine
which weather data affect

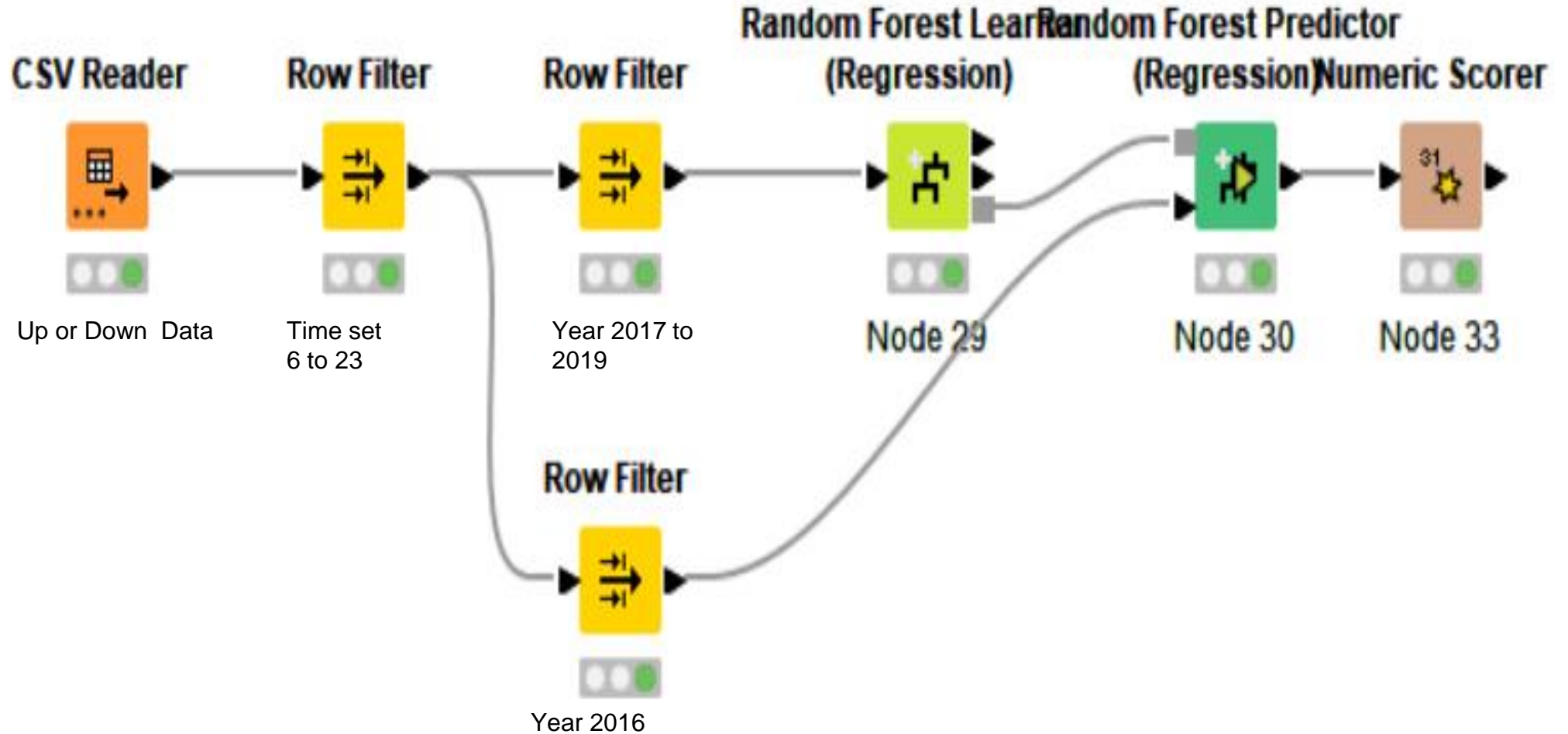
Sensible_Temp + Humidity is best !!!

If Wind is included, the R square value
is too low when MLR is analyzed

Part 3

Random Forest

Random Forest is an algorithm that can predict using various factors while avoiding overfitting



Part 3

Random Forest

Rain
Wind
Humidity
Sensible_temp
Hour
Rain(hour)
Wind(hour)
Humidity(hour)
Sensible_temp(hour)
Usage(hour)
Rain(week)
Wind(week)
Humidity(week)
Sensible_temp(week)
Usage(week)

Feature:15

Random Forest Learner
(Regression)



Statistics - 4:3...	
File	
R ² :	0,943
Mean absolute error:	51,117
Mean squared error:	6,401,867
Root mean squared error:	80,012
Mean signed difference:	-10,094
Mean absolute percentage error:	0,113
Adjusted R ² :	0,943

UP data

Tree depth: 10

Number of models: 500

Part 3

Random Forest

Row ID	#splits (level 0)	#splits (level 1)	#splits (level 2)	#candidates (level 0)	#candidates (level 1)	#candidates (level 2)
Rain	1 ✓	9	20	107	198	419
Wind	16	35	69	91	179	393
Humidity	27	40	91	101	181	388
Sensible_te...	17	54	135	104	181	398
Hour	90	158	319	102	197	400
Rain(hour)	1 ✓	9	17	97	226	417
Wind(hour)	50	45	72	110	219	383
Humidity(h...	38	49	91	91	202	404
Sensible_te...	6	68	151	96	183	419
Usage(hour)	71	168	298	95	229	412
Rain(week)	0 ✓	0	11	105	193	399
Wind(week)	37	47	71	103	198	374
Humidity(w...	66	73	124	109	216	393
Sensible_te...	2 ✓	36	138	111	185	386
Usage(week)	78	209	393	78	213	415

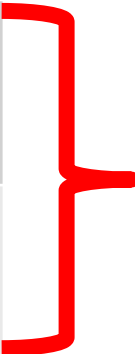

Part 3

Random Forest

Row ID	#splits (level 0)	#splits (level 1)	#splits (level 2)	#candidates (level 0)	#candidates (level 1)	#candidates (level 2)
Rain	0 ✓	8	17	98	190	380
Wind	31	40	78	105	205	417
Humidity	7	32	92	101	200	410
Sensible_te...	16	57	119	98	208	390
Hour	71	161	332	84	192	407
Rain(hour)	0 ✓	9	25	96	197	413
Wind(hour)	59	78	123	99	207	388
Humidity(h...	35	70	146	97	196	397
Sensible_te...	37	97	148	104	216	404
Usage(hour)	83	131	297	106	194	403
Rain(week)	0 ✓	0	3	113	215	397
Wind(week)	56	65	98	100	194	374
Humidity(w...	2 ✓	11	54	107	203	415
Sensible_te...	4	33	76	93	175	390
Usage(week)	99	208	391	99	208	415

Part 3

Random Forest

Number of variable	R-Squared value	
15	0.932	 UP data Tree depth: 8 Number of models: 500
11	0.943	
15	0.925	 DOWN data Tree depth: 8 Number of models: 500
13	0.938	

Part 3

Random Forest

Evaluation index	Multiple Linear Regression(MLR)		Random Forest
R^2	0.891	0.047	0.938
MAE	66.009	18.868	47.141
MSE	8970.951	3846.127	5124.824
RMSE	94.715	23.127	71.588

A 3D scene with a dark blue background and a dark grey floor. In the center, there is a large, glowing orange cube. To its left and slightly behind, there is a smaller, translucent blue cube. Below the large orange cube, there is a dark grey rectangular prism. To the right of the large orange cube, there is a small black cylinder. Two bright yellow circles are positioned around the text: one to the left and one above and to the right. A small, stylized carrot icon is located to the right of the text.

Thank you