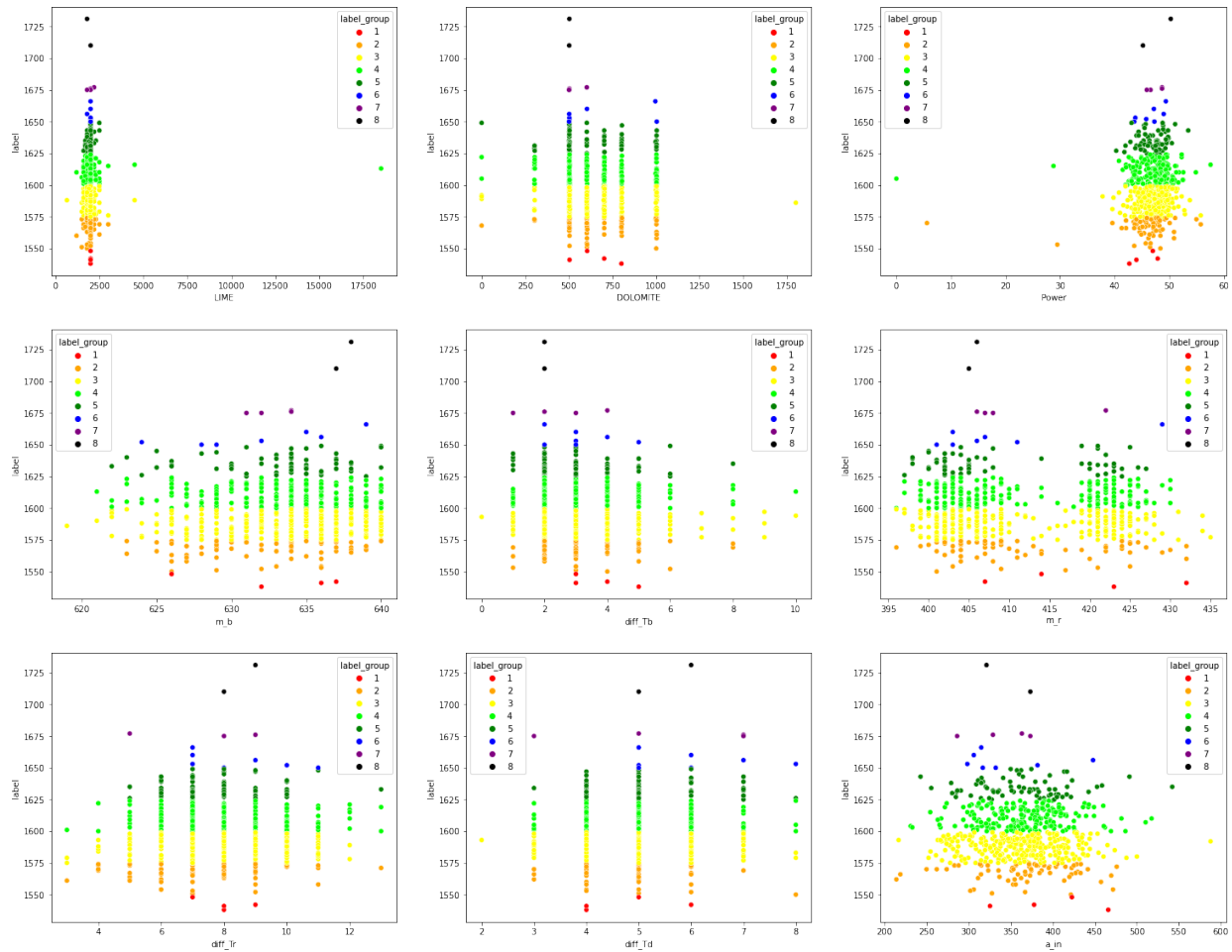


출강온도 예측 오차 분석 결과

RandomForest 모델의 출강온도 예측 성능 : **RMSE 23**

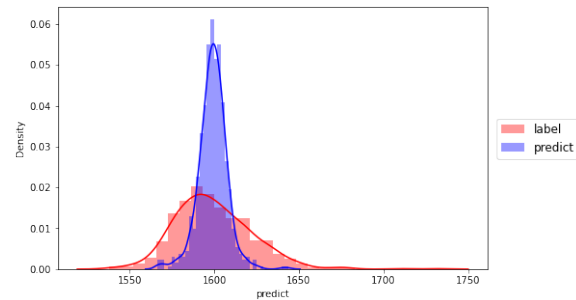


먼저 출강온도와 각 변수 간 산점도를 그렸습니다. 1~8로 출강온도를 아래의 기준으로 그룹화했습니다.

- min ~ 1550 : 1
- 1550 ~ 1575 : 2
- 1575 ~ 1600 : 3
- 1600 ~ 1625 : 4
- 1625 ~ 1650 : 5
- 1650 ~ 1675 : 6
- 1675 ~ 1700 : 7
- 1700 ~ max : 8

위 산점도를 보시면, 각 변수와 출강온도 간의 상관관계가 크게 없다는 것을 확인할 수 있었습니다.

모델의 예측값과 실제 출강 온도의 차이를 prediction_diff라는 새로운 칼럼에 추가하여 분석해본 결과, 모델이 대부분의 데이터에 대해 1600도 근처로 예측하고 있음을 확인했습니다. 아래 이미지는 예측값(파란색)과 실제 출강온도(빨간색)의 분포 그래프입니다.



실제로 아래 결과 이미지를 보면, 대부분 predict 값이 1600 전후인 것을 확인하실 수 있습니다.

```
In [118]: result[result["label"]<1575].sort_values(by="label")
Out[118]:
```

	HEAT	LIME	DOLOMITE	Power	m_b	dH_Tb	m_r	dH_Tr	dH_Td	a_in	label	predict	prediction_diff	label_group
11	473851	2000.0	802.0	42.7	630	5	423	8	4	466.12054	1538	1595.87	57.87	1
138	473839	2001.0	504.0	44.0	630	3	432	8	4	325.277894	1541	1567.17	26.17	1
12	474032	2000.0	703.0	47.0	627	4	407	9	6	377.70350	1542	1599.54	57.54	1
49	474041	2002.0	605.0	47.0	626	3	414	7	5	422.937727	1548	1586.50	38.50	1
17	474763	1802.0	1002.0	48.4	626	4	401	7	8	421.640664	1550	1604.21	54.21	2
23	473835	1501.0	654.0	46.5	629	3	417	7	5	327.650111	1551	1598.30	47.30	2
70	474832	2004.0	504.0	45.5	630	6	409	9	6	367.086721	1552	1587.56	35.56	2
40	473731	1804.0	600.0	39.5	630	1	421	7	4	301.803225	1553	1593.86	40.86	2
36	474056	1802.0	600.0	45.7	630	3	426	6	4	356.004440	1554	1596.16	42.16	2
35	474076	2005.0	604.0	46.6	630	3	403	6	5	471.758827	1554	1596.46	42.46	2
53	474327	2003.0	694.0	48.2	626	2	408	9	4	355.819126	1556	1596.47	37.47	2
27	474445	2003.0	1004.0	50.9	627	5	404	11	6	352.853966	1558	1603.50	45.50	2
74	474344	2002.0	703.0	46.4	629	2	406	5	5	413.639723	1561	1595.86	34.86	2
158	474354	2004.0	1002.0	44.9	627	2	411	3	5	345.029047	1561	1595.30	34.30	2
34	474299	2001.0	1003.0	45.4	631	3	404	6	5	354.633091	1562	1595.21	43.21	2
13	474676	2004.0	603.0	47.0	626	1	404	5	3	313.787691	1562	1618.02	57.02	2
61	474445	2003.0	1004.0	50.9	627	2	402	8	5	311.046219	1563	1599.58	36.58	2
68	474109	1802.0	803.0	44.6	630	3	412	9	5	362.342476	1563	1598.70	35.70	2
188	474271	2002.0	502.0	47.6	626	3	414	6	5	419.869688	1564	1593.93	21.93	2
91	474086	2003.0	504.0	48.4	627	4	407	6	5	356.456669	1564	1598.58	44.58	2
81	473985	2002.0	694.0	47.3	626	4	405	7	4	373.917058	1565	1596.68	31.68	2

실제 출강온도 값이 1575 이하인 경우입니다. 1600과 비교적 차이가 있으므로, 예측값과의 차이인 prediction_diff 역시 높습니다.

```
result[(result["label"]>1595) & (result["label"]<1605)].sort_values(by="label")
```

	HEAT	LIME	DOLOMITE	Power	m_b	dH_Tb	m_r	dH_Tr	dH_Td	a_in	label	predict	prediction_diff	label_group
550	473483	2002.0	1000.0	48.8	639	3	428	8	6	301.803225	1596	1599.09	3.09	3
584	474240	1801.0	803.0	48.1	626	3	387	7	4	282.584715	1596	1593.36	2.64	3
688	474320	2003.0	502.0	49.7	632	2	399	11	6	320.911787	1596	1598.06	2.06	2
555	474146	2001.0	382.0	48.0	629	3	409	9	6	303.803886	1596	1599.29	3.29	3
594	474658	2003.0	503.0	49.5	638	4	406	9	6	370.348406	1596	1597.75	1.75	3
604	474673	2004.0	503.0	49.5	633	5	402	9	5	376.576241	1596	1601.78	5.78	3
519	474781	2005.0	1005.0	48.7	632	2	403	6	5	338.910749	1596	1600.92	4.92	3
616	473837	2003.0	503.0	48.4	635	1	421	8	4	295.620380	1596	1598.45	2.45	3
485	473960	1702.0	701.0	46.7	630	2	425	5	3	349.890302	1596	1603.14	7.14	3
497	473809	2002.0	502.0	44.8	632	7	419	10	6	429.355679	1596	1590.52	5.48	3
315	473631	2003.0	506.0	48.5	632	3	423	10	5	372.424018	1596	1593.14	15.86	3
487	474986	2005.0	505.0	50.3	640	3	404	7	5	435.587912	1597	1603.30	6.30	3
371	474332	1902.0	604.0	48.8	633	2	404	7	7	275.483306	1597	1606.47	12.47	3
673	474608	2005.0	704.0	51.2	634	2	406	7	6	287.027428	1597	1594.46	2.54	3
532	473917	1804.0	504.0	44.6	635	2	423	6	4	340.884972	1597	1601.25	1.25	3
620	473953	2003.0	603.0	47.8	626	9	421	7	4	424.794205	1597	1593.57	5.43	3
305	473514	2001.0	502.0	51.5	640	3	431	9	5	381.618212	1597	1590.79	16.21	3
630	473914	2003.0	703.0	44.4	632	4	421	7	5	301.556708	1597	1592.14	4.86	3
625	473586	1903.0	705.0	50.6	637	2	424	6	4	375.389177	1597	1597.20	0.20	3
658	473710	1702.0	803.0	43.9	636	3	421	7	6	386.140209	1598	1594.86	3.14	3
385	474098	2003.0	704.0	50.7	635	4	404	5	5	375.690590	1598	1595.11	2.89	3
229	473617	2004.0	382.0	48.1	639	4	423	8	4	405.465771	1598	1615.36	20.36	3
467	474005	2002.0	703.0	48.0	632	2	417	6	6	389.890587	1598	1605.80	7.80	3

실제 출강온도 값이 1595~1605인 경우입니다. 비교적 prediction_diff가 낮은 것을 확인할 수 있습니다.

```
result[result["label"]>1650].sort_values(by="label")
```

	HEAT	LIME	DOLOMITE	Power	m_b	dH_Tb	m_r	dH_Tr	dH_Td	a_in	label	predict	prediction_diff	label_group
7	474091	2005.0	504.0	45.8	624	5	411	10	5	381.912528	1602	1583.63	68.37	6
10	474696	2005.0	503.0	43.8	632	3	406	7	8	296.296233	1603	1592.22	65.78	6
14	474652	1803.0	502.0	46.0	636	4	407	9	7	448.026882	1606	1598.20	56.80	6
16	474698	2002.0	604.0	47.2	635	3	403	7	6	306.024448	1600	1604.54	55.46	6
4	473482	2002.0	987.0	49.4	639	2	429	7	5	314.889826	1606	1590.51	75.49	6
6	474160	2001.0	503.0	46.6	631	1	408	8	3	373.313586	1605	1605.39	69.61	7
4	474641	1803.0	501.0	45.9	632	3	407	8	7	286.137880	1605	1600.67	65.33	7
9	474623	2003.0	503.0	48.7	634	2	406	9	7	328.836175	1606	1600.17	75.83	7
2	473648	2202.0	603.0	48.7	634	4	422	5	5	383.230204	1607	1594.66	82.34	7
1	474672	2004.0	502.0	45.2	637	2	405	6	5	373.313266	1710	1601.52	108.48	8
8	474642	1803.0	503.0	50.3	638	2	406	9	6	321.026761	1721	1608.71	121.29	8

실제 출강온도 값이 1650 이상인 경우입니다. 높은 prediction_diff 인 것을 확인할 수 있습니다.