Dataset for temperature

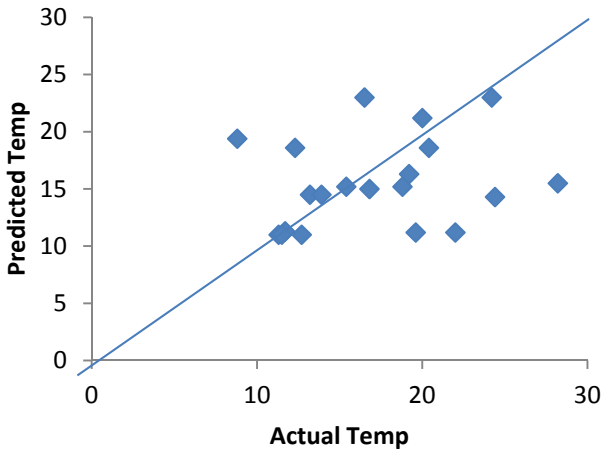
1st column is the actual temperature, 2rd column is the Google predicted temperature.

Temp	Pred_temp	Cloud3pm	WindSpeed3pm	Pressure3pm	Date
20	21.2	8	6	1015.6	12/19/2007
20.4	18.6	7	13	1014.1	12/20/2007
28.2	15.5	7	22	1008.3	12/21/2007
19.6	11.2	3	30	996.8	12/22/2007
18.8	15.2	7	20	1009.1	12/23/2007
8.8	19.4	6	22	1008.1	4/28/2008
13.2	14.5	1	9	1019.6	4/29/2008
13.9	14.5	7	19	1020.5	4/30/2008
15.4	15.2	7	26	1013.8	5/6/2008
16.5	23	1	30	1013.5	5/7/2008
12.3	18.6	7	15	1022.1	6/23/2008
11.7	11.3	6	28	1024.2	6/24/2008
11.3	11	1	28	1021.9	6/25/2008
11.5	11	1	28	1020.4	6/26/2008
12.7	11	1	24	1025	6/27/2008
22	11.2	3	28	1006.4	12/15/2008
24.2	23	1	31	1009.4	12/16/2008
16.8	15	7	30	1008.6	12/17/2008
24.4	14.3	4	33	1001.1	12/18/2008
19.2	16.3	2	15	1006.7	12/19/2008

Mean of expected value is 17.48, the predicted is 16.39. And corresponding standard deviation is 5.25 and 5.46 respectively. Since we do not know the distribution of temperature, therefore, we have two choices to compare the actual value and predicted one. First, according to central limit theory, we can assume it follows normal to approximate the data. The other is non-parametric method. We conducted two-sample Kolmogorov-Smirnov good of fit test for actual temperature and predicted temperature. D=0.25, and p-value is 0.56. Therefore, we cannot reject the H0 at alpha=0.05 level. Thus, we think it is a good fit between the actual temperature and predicted value.

If we assume the distribution follows normal, and we conduct the chi-square good of fit test. Chi-square=1.969785, DOF=17, pvalue=0.99. This result also tell us that over- all- fit is good.

In order to check the fit, we plot the actual temperature with predicted one. We find there are 6 points away from the matching line, which indicates some problems of prediction.



Conclusion: overall fitting is good, but still need to improve.