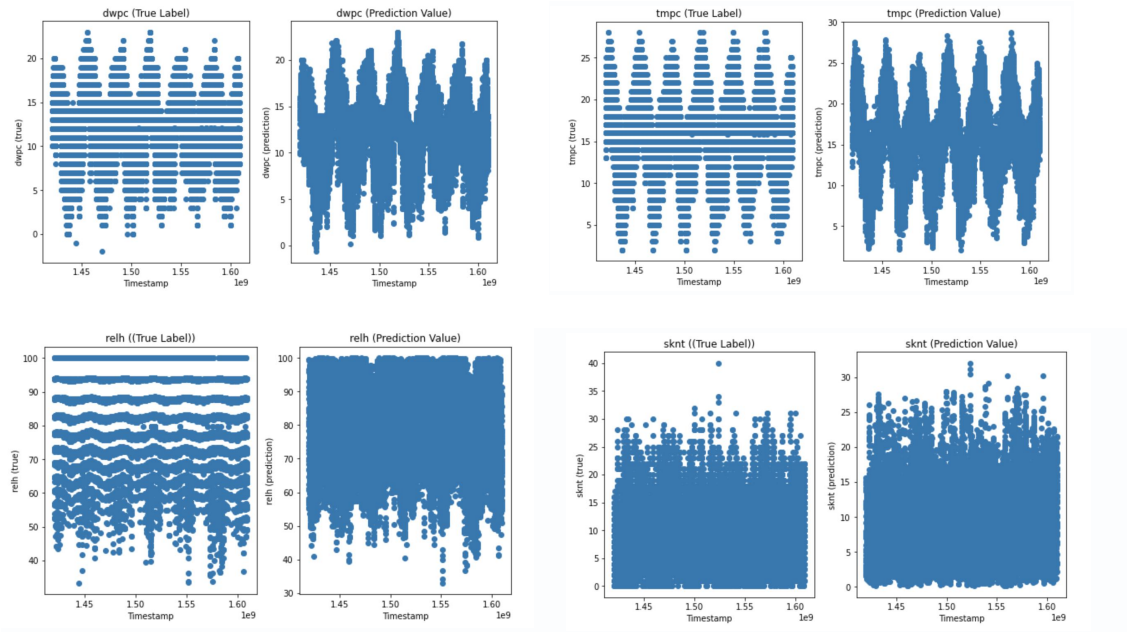
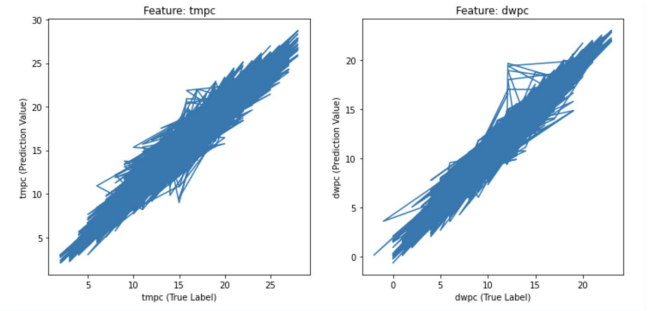
Model Evaluation

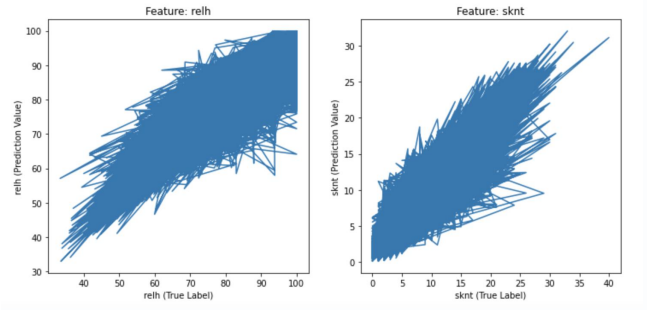
After fitting the model, the model accuracy can be measured by score(X_test, Y_test), which returns the coefficient of determination of the prediction. Our model gets the result of 0.91, which is close to the best possible score of 1. Therefore, the overall accuracy of our model prediction is very high.

In addition to that, we also check the prediction accuracy of each weather feature. Based on the scatter plot, we find a high distribution similarity between our prediction results and the actual values. Especially for features 'dwpc' and tmpc’. Therefore, our model can correctly predict the trends of four weather features. The prediction of 'dwpc' and tmpc’ is especially accurate.



The following plot results show us a similar outcome. The comparison between the predicted results and the actual values of 'dwpc' and tmpc’ is closer to a y=x straight line. At the same time, the prediction results of the other two features ('relh' and 'sknt') obviously have some deviations.





We also use the MSE function to measure the mean square error of the prediction result of each feature. We found that the MSE of features 'dwpc' and tmpc' is low. The MSE of feature 'relh' is much larger than the other features.

```
# Measureing the mean square error

from sklearn.metrics import mean_squared_error
MSE1 = mean_squared_error(Y_output[['tmpc']],Y_test[['tmpc']])
MSE2 = mean_squared_error(Y_output[['dwpc']],Y_test[['dwpc']])
MSE3 = mean_squared_error(Y_output[['relh']],Y_test[['relh']])
MSE4 = mean_squared_error(Y_output[['sknt']],Y_test[['sknt']])
print("The MSE of feature tmpc is: ", MSE1)
print("The MSE of feature dwpc is: ", MSE2)
print("The MSE of feature relh is: ", MSE3)
print("The MSE of feature sknt is: ", MSE4)
```

```
The MSE of feature tmpc is:  0.5203489162568209
The MSE of feature dwpc is:  0.37716528852742087
The MSE of feature relh is:  24.738135195360936
The MSE of feature sknt is:  4.046848024673869
```

Overall, Our prediction model performs very well in terms of accuracy. The prediction for ‘tmpc’ （Temperature of the environment in celsius） and 'dwpc' (Temperature of the dew point in the environment in celsius) is highly accurate. In contrast, the prediction for feature ‘relh’ (Relative humidity of the environment in percentage) is relatively poor.