Air quality forcasting:

In this project we should predict PSI levels. I do this prediction for each of the locations individualy(‘central’, ‘east’, …).

For time forecasting, we just have sequentialy data. So we should extract useful features for regression analysis. For this work, we use **forcasting_features** function, which select **n_days** interval before current day, as a common features. For example, if we set **n_days** as 5, number of features of our dataset would be 5.

After we construct our dataset, we should select our regression algorithm for predicting psi levels of 500 days after our choosen **start_day**. Although, we used 2000 days before **start_day** for train set and train our model for this set.

I used two regression method:

1) SVR (Support Vector Regressor) – regression version of svm

2) KNeighborsRegressor - Regrerssion version of KNN

This two method are implemented by sickit-learn.

RESULTS :

----

* For each of the methods, I plot **PSI-**variation diagram value of 500-days and its prediction. This diagrams show that prediction of our model are very similar and models accuarate very well.

Differences between two models:

1) SVR has a longer training time.
   But Knn has a short training time

2) SVR digram shows that it’s very similar two original **PSI values**
   Knn is accurate too, but SVR is better.

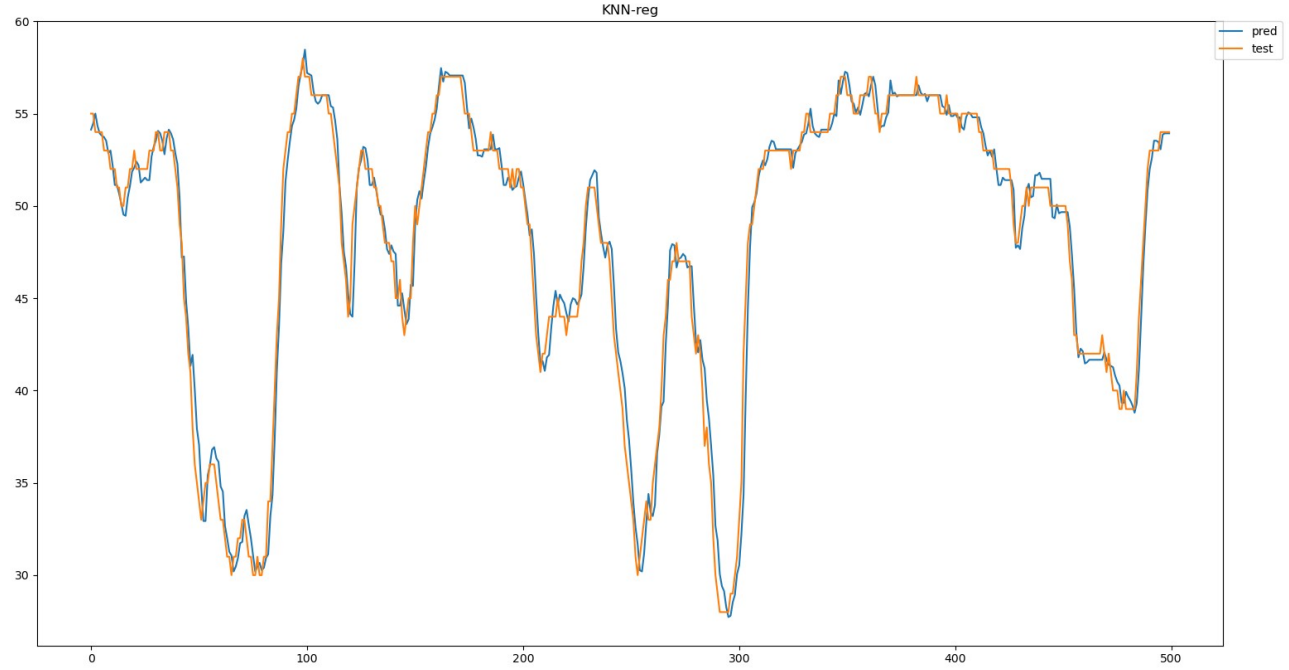
3) Score prediction of SVR is better than KNN

* In my experiment, I used ‘central’ location for psi prediction. We can do that for other location-features individualy.

* Another work that we can do that, is combining all locations together. Then forecast their values. Bu this method is accurate very well

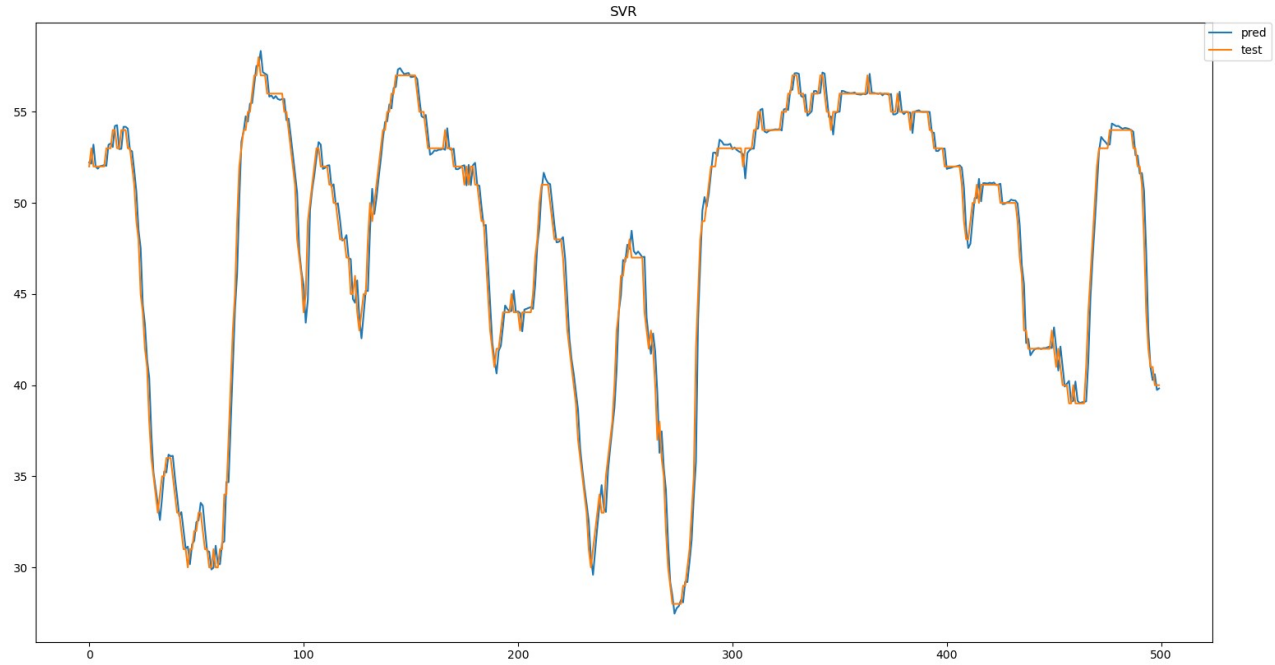
* I used just 2000 days as training set and predict 500 days after that. So for any intervals in the dataset, between 2000 and 30000, we can predict our values.

KNN-reg Digram for ‘central’ location:



Score : 0.9586

SVR Diagram for ‘central’ location:



Score : 0.9811

Thanks.