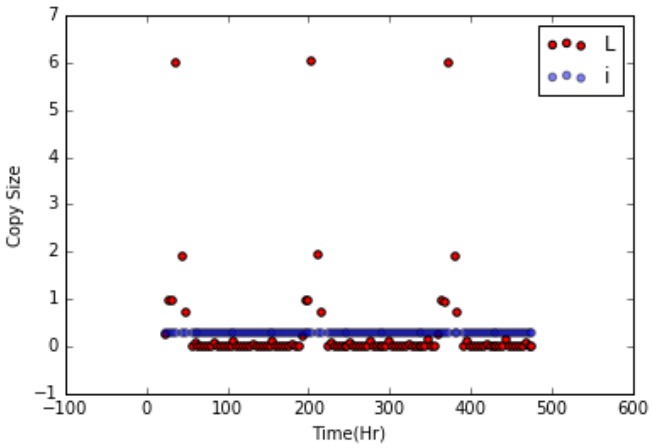
RESULT: NETWORK BACKUP DATASET

1. The indexes of the Significant variables(32) in the model(one-hot encoded) are as :

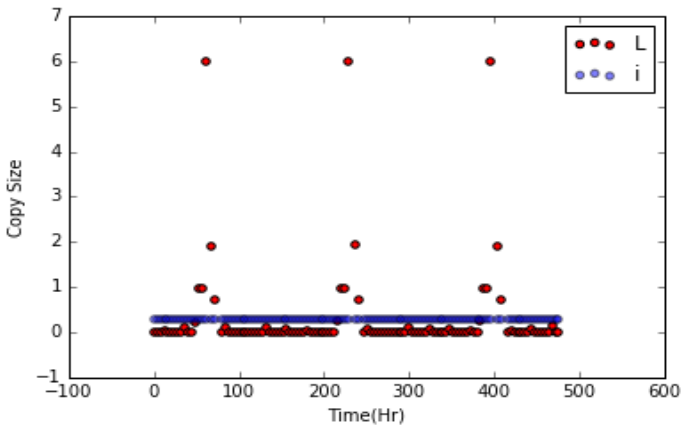
   Indexes of selected features:

   [28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 48, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62]
2. RMSE obtained is as: *0.0903*
3. Fitted values and Actual values scattered plot over time:
   - On evaluating the plot, actual and predicted copy sizes of all the files for a given workflow over time, we get a pattern. In each period, majority of the time, the file copy size is similar and then it peaks for a small time interval before the next cycle begins. However, the predicted copy size is found to be similar.
   - Figures below show the plots of copy sizes of all the files (actual in Red and predicted in blue) plotted against the time (we have added the copy sizes for all the files at the particular time) for *Workflow -1.*

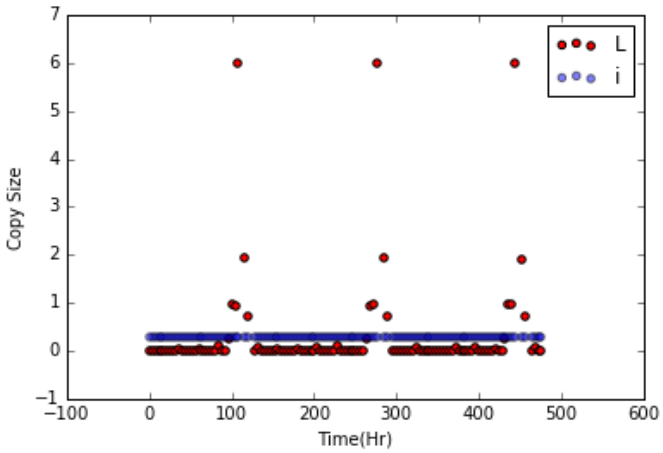
Workflow-1 Days 1-20



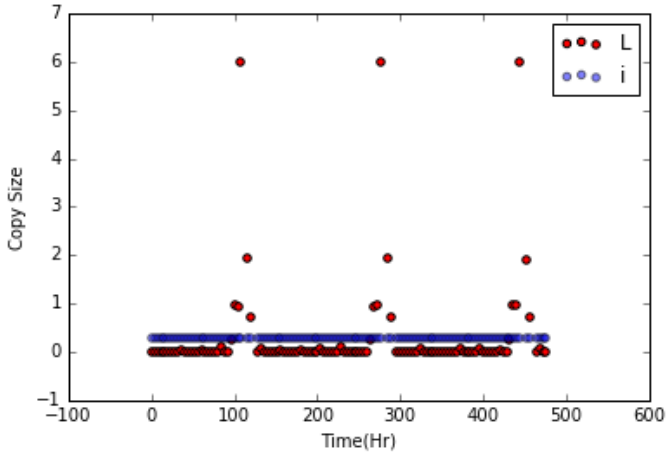
Workflow-1 Days 21-40



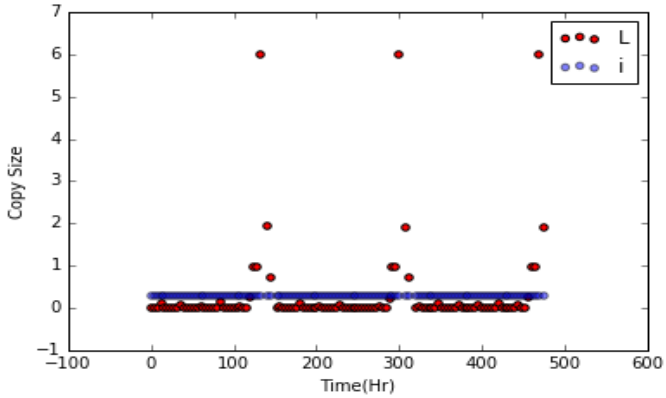
Workflow-1 Days 41-60



Workflow-1 Days 61-80

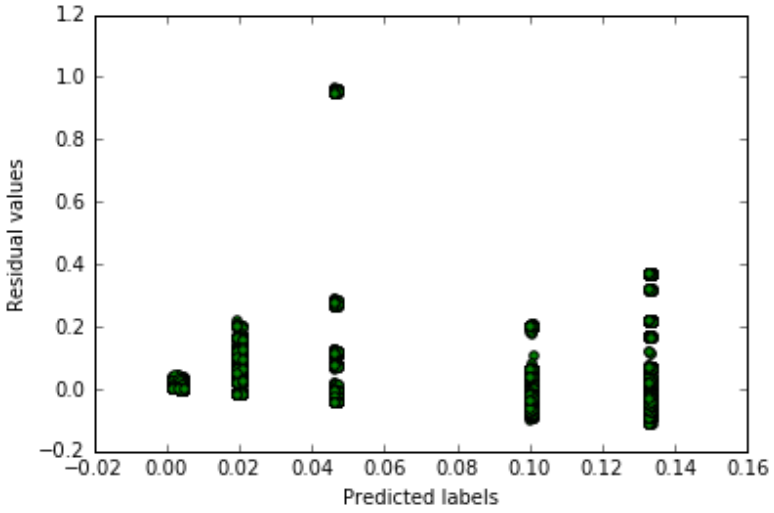


Workflow-1 Days 81-100



- The other Workflows were found to show a *similar pattern*.

4. The Residuals versus Fitted values plot is as:



RESULT: BOSTON HOUSING DATASET

1. The Significant variables(6) obtained are:
    - CHAS: Charles River dummy variable.
    - NOX: nitric oxides concentration (parts per 10 million).
    - RM: average number of rooms per dwelling.
    - DIS: weighted distances to five Boston employment centers.
    - PTRATIO: pupil-teacher ratio by town.
    - LSTAT: % lower status of the population
2. RMSE obtained is as : *5.0378*
3. The Residuals versus Fitted values plot is as: