Time Series Forecasting Using N-BEATS and RevIN

Introduction

N-BEATS focuses on the univariate times series point forecasting problem using deep learning. This deep neural architecture is based on backward and forward residual links and a very deep stack of fully-connected layers. It is interpretable, applicable without modification to many target domains, and fast to train, making it widely usable.

Although different kinds of architecture are widely used for time series forecasting, namely Statistical Model, Hybrid Model, and Purely Deep Learning models, all the models ignore the distribution shift problem. Statistical properties such as mean and variance often change over time in time series, i.e., time-series data suffer from a distribution shift problem. To overcome this issue, RevIN provides a simple yet effective normalization method called reversible instance normalization (RevIN), a generally-applicable normalization-and-denormalization method with learnable affine transformation added on symmetric layers.

Why Does Revin Work?

In statistics and applications of statistics, normalization or standardization is a common practice for the statistician. We know from **Central Limit Theorem (CLT)** in probability theory that in many situations, when independent random variables are summed up, their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed or specifically standard normal distribution. If X_1, X_2,..., X_n,... are random samples drawn from a population with overall mean μ and finite variance σ, and if the sample mean is x then the limiting distribution of $(x-\mu)/\sigma \sim N(0,1)$ standard normal distribution.

So, RevIN is doing a similar thing. It is doing a similar operation on the data point of every timestamp that makes the data distribution follow an approximate standard normal distribution. Standard Normal distribution is well-defined and has a defined range. So, RevIN is taking the train and test distribution on the same scale. So, it is comparable and this is making the model give a more accurate forecast.

Experiments

Given the 50 time-series datasets, I have modeled the forecasting model with N-BEATS. I have used the Generic DL version of N-BEATS as interpretability is not the task here and to overcome the distribution shift I have used RevIN. I have used the first 75% of the data as training data and the last 25% for the evaluation.

Implementation Details

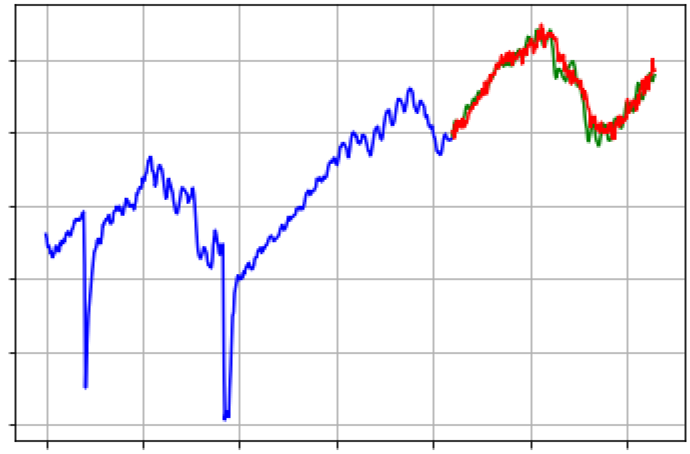
I have used the Generic N-BEATS model to forecast the last 25% of the dataset. So the forecast length H is 0.25 * the length of data. As input window to a multiple of the forecast horizon H, and typical lengths of x in the paper is from 2H to 7H, we have used a backcast length is 2 * forecast length. Setting backcast length more than 2H will not yield that much data to train.

RevIN can handle multivariate time-series data. Here, I need to forecast the univariate time series. So, I have used hyperparameter num_features =1 to handle univariate data. Apart from this, I have used MSE as a loss function and Adam optimizer and a batch size of 10, and the learning rate lies in the range 1e-3 to 1e-4.

Plot and Benchmark

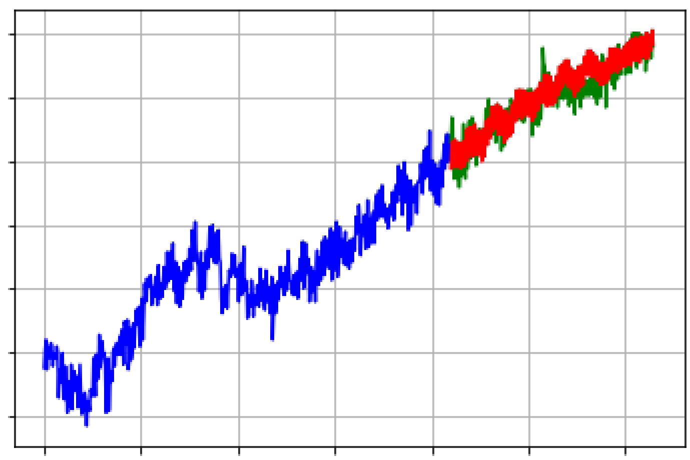
Dataset 0

mae : 0.022 mse : 0.001 rmse : 0.030 mape : 0.045



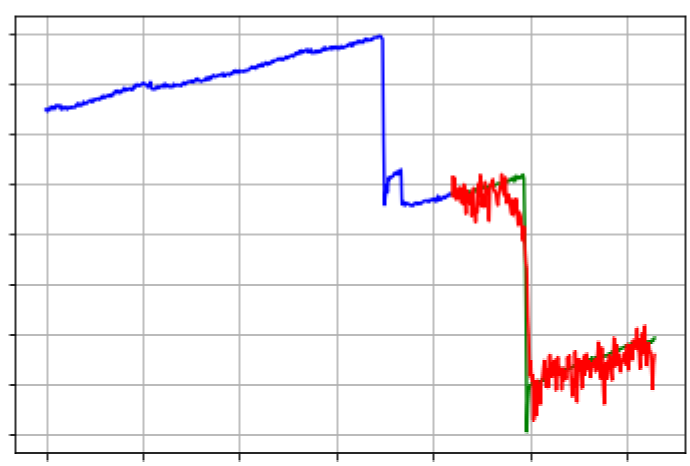
Dataset 1

mae : 0.022 mse : 0.001 rmse : 0.030 mape : 0.045



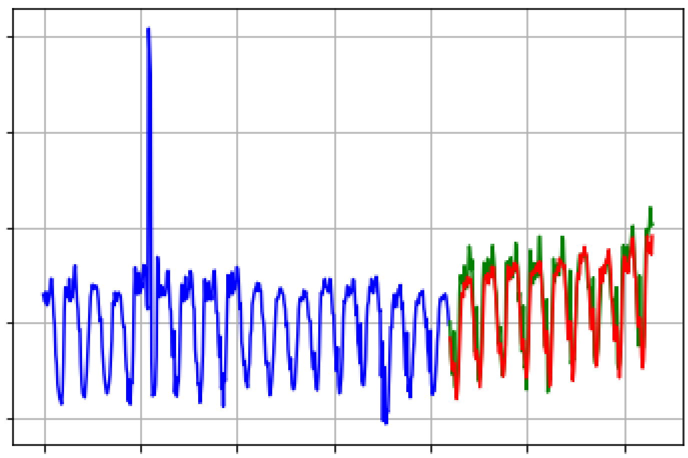
Dataset 2

mae : 0.145 mse : 0.049 rmse : 0.220 mape : 0.326



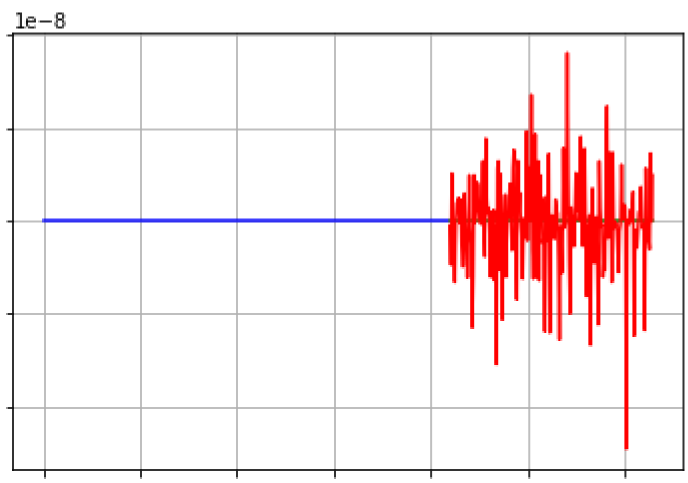
Dataset 3

mae : 27.839 mse : 1374.780 rmse : 37.078 mape : 0.812



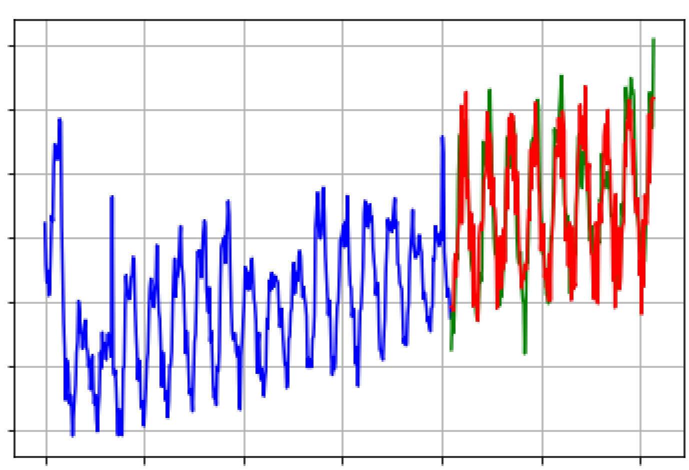
Dataset 4

mae : 0.000 mse : 0.000 rmse : 0.000 mape : nan



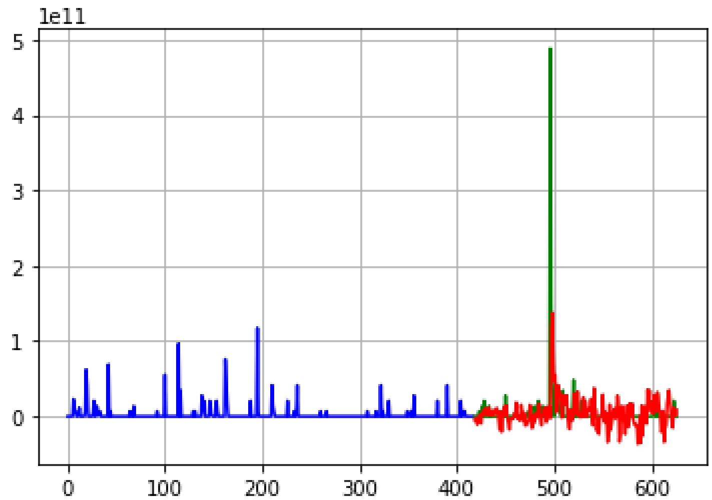
Dataset 5

mae : 0.025 mse : 0.001 rmse : 0.030 mape : 0.587

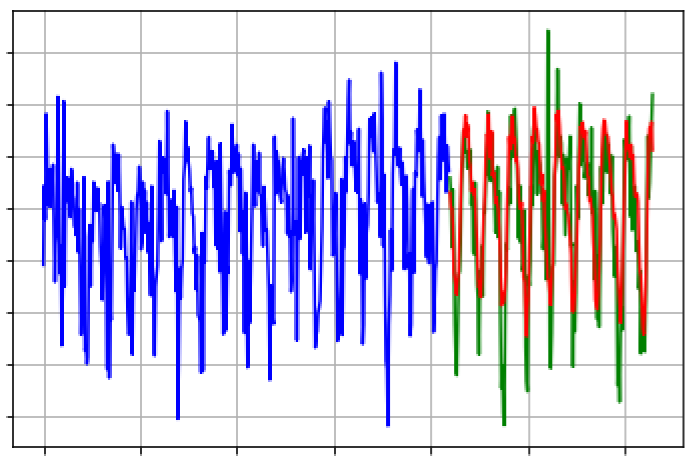


Dataset 6

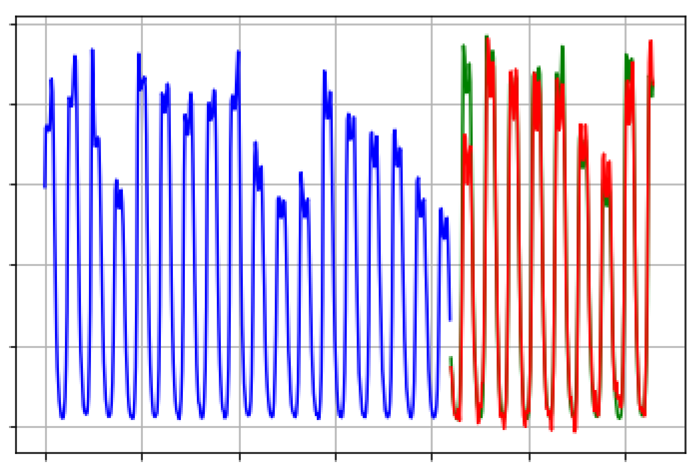
mae : 10336333730.903 mse : 1330740867279751217152.000 rmse : 36479321091.267 mape : 240.607



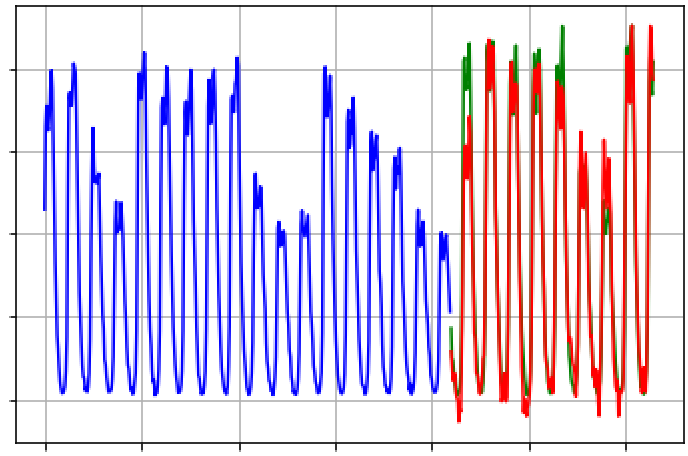
Dataset 7 mae : 133.740 mse : 31215.635 rmse : 176.679 mape : 8.873



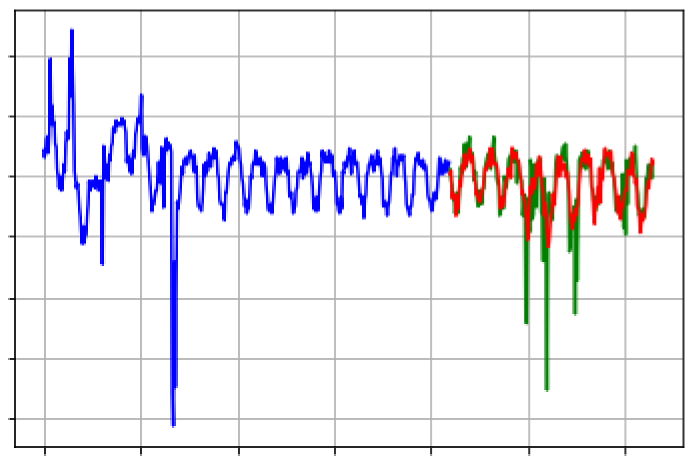
Dataset 8 mae : 1031.376 mse : 2568970.556 rmse : 1602.801 mape : 37.839



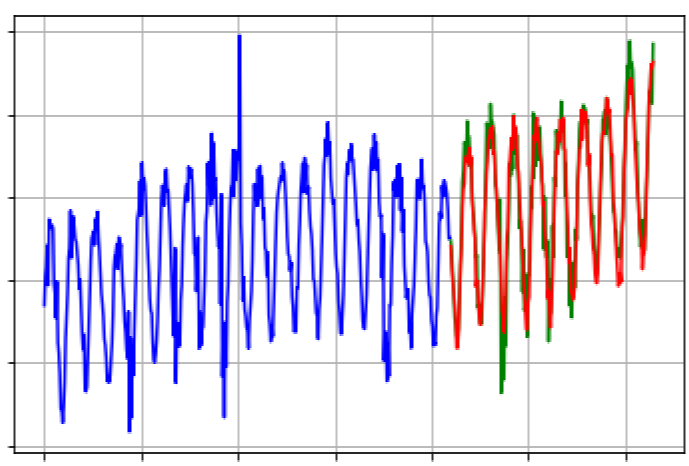
Dataset 9 mae : 2401.873 mse : 11268494.172 rmse : 3356.858 mape : 87.691



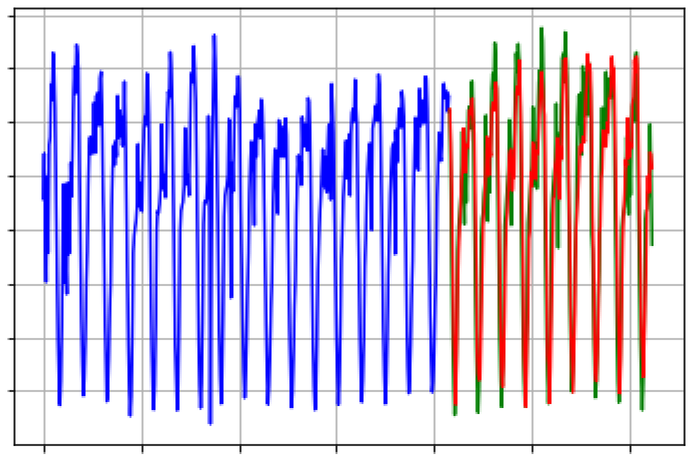
Dataset 10 mae : 0.063 mse : 0.010 rmse : 0.101 mape : 0.910



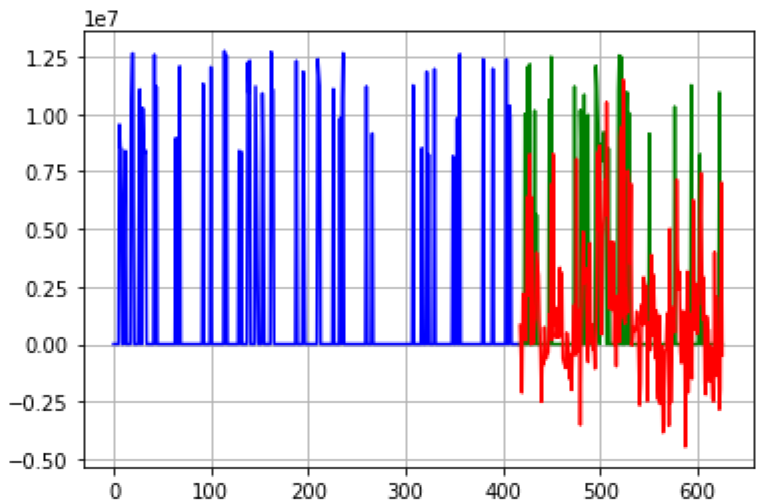
Dataset 11 mae : 0.022 mse : 0.001 rmse : 0.029 mape : 0.243



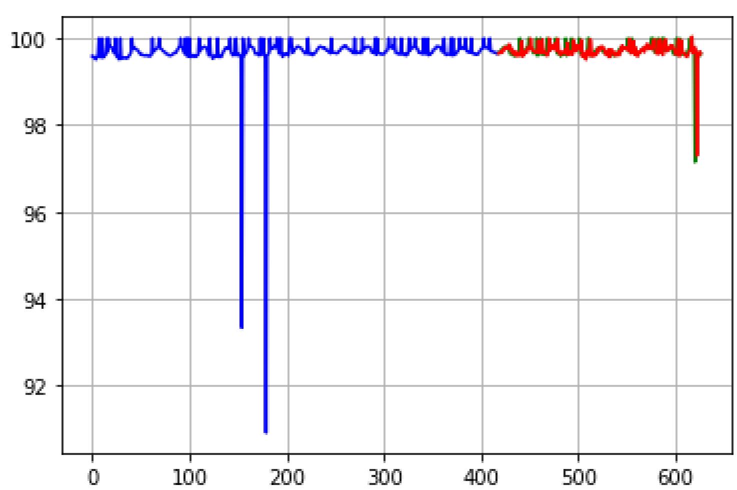
Dataset 12 mae : 3418.636 mse : 17287434.407 rmse : 4157.816 mape : 14.251



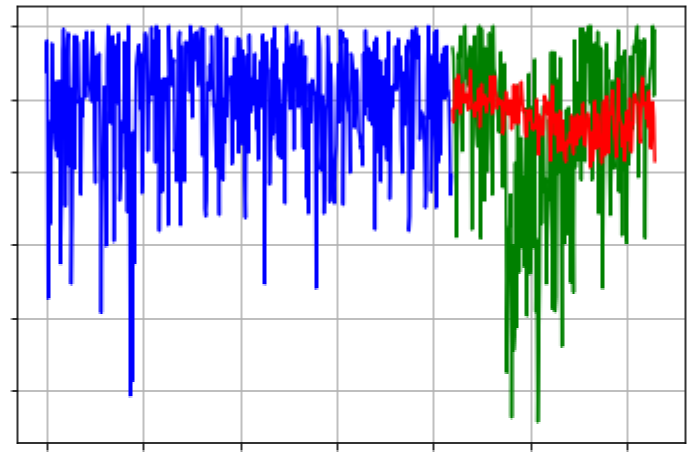
Dataset 13 mae : 2728964.613 mse : 15925176250174.605 rmse : 3990636.071 mape : 1422.792



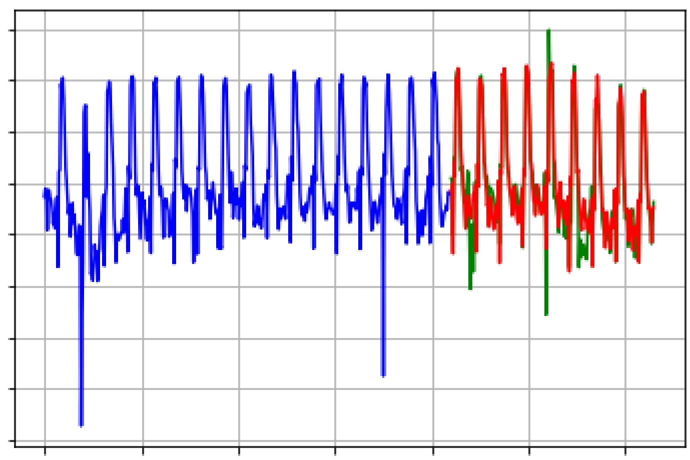
Dataset 14 mae : 0.110 mse : 0.046 rmse : 0.214 mape : 0.110



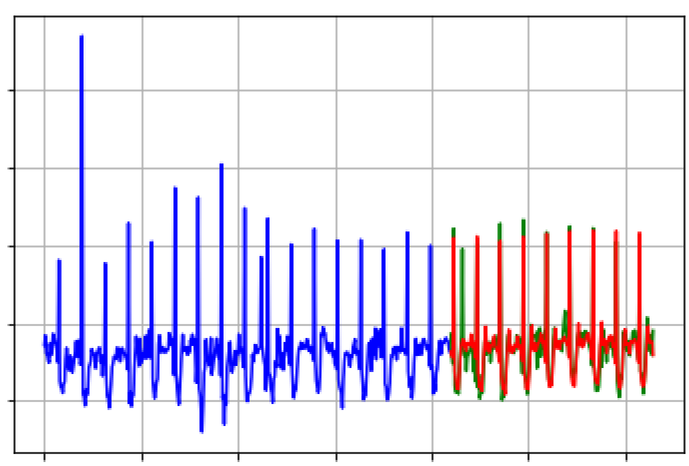
Dataset 15 mae : 0.488 mse : 0.390 rmse : 0.625 mape : 0.491



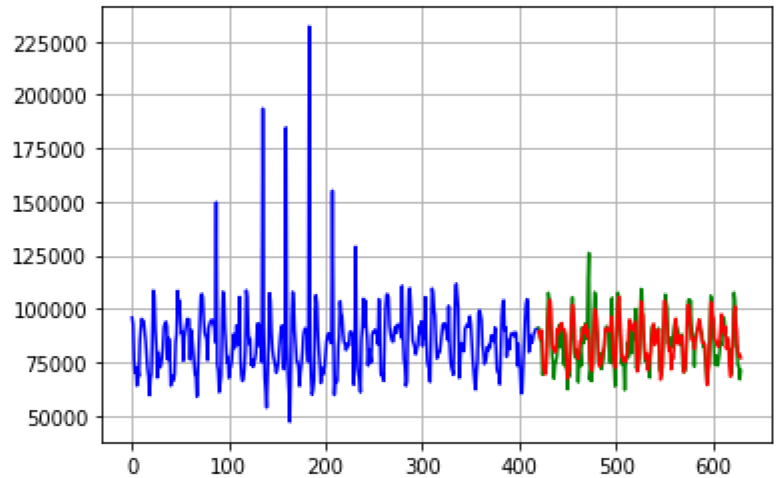
Dataset 16 mae : 0.256 mse : 0.131 rmse : 0.362 mape : 0.281



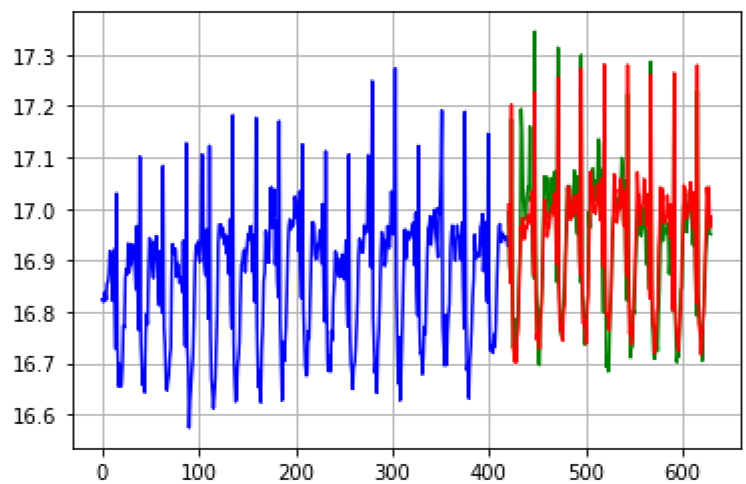
Dataset 17 mae : 0.055 mse : 0.006 rmse : 0.078 mape : 1.143



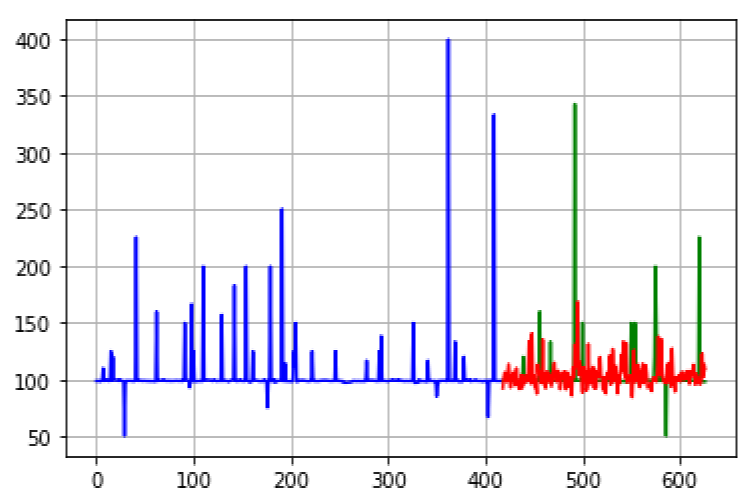
Dataset 18 mae : 4901.860 mse : 46128883.058 rmse : 6791.825 mape : 5.723



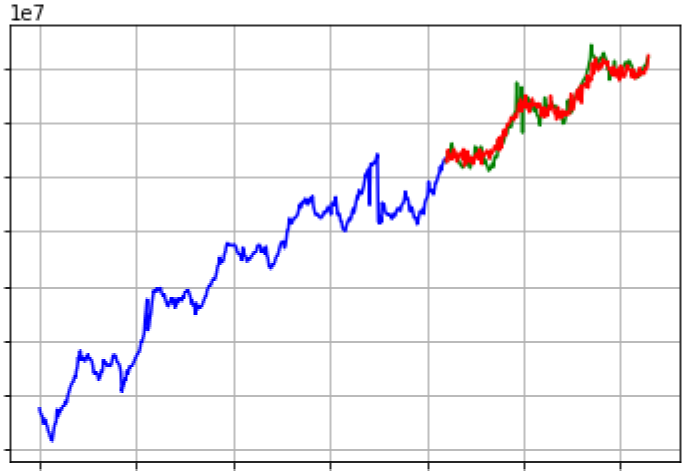
Dataset 19 mae : 0.038 mse : 0.003 rmse : 0.052 mape : 0.227



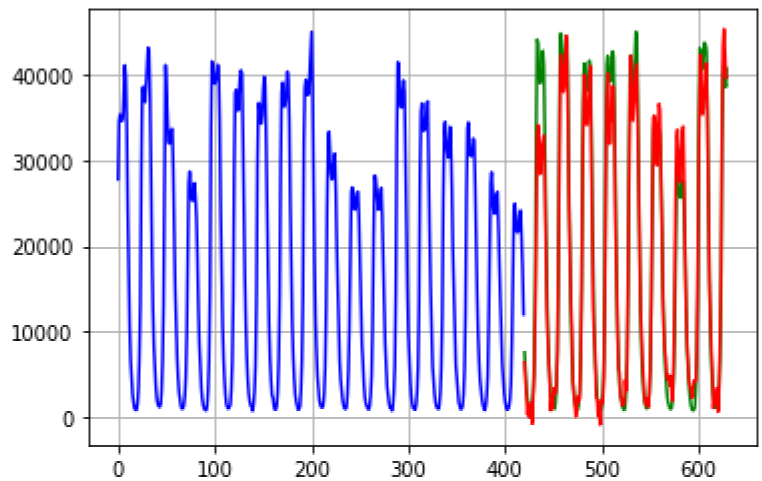
Dataset 20 mae : 11.960 mse : 553.997 rmse : 23.537 mape : 11.159



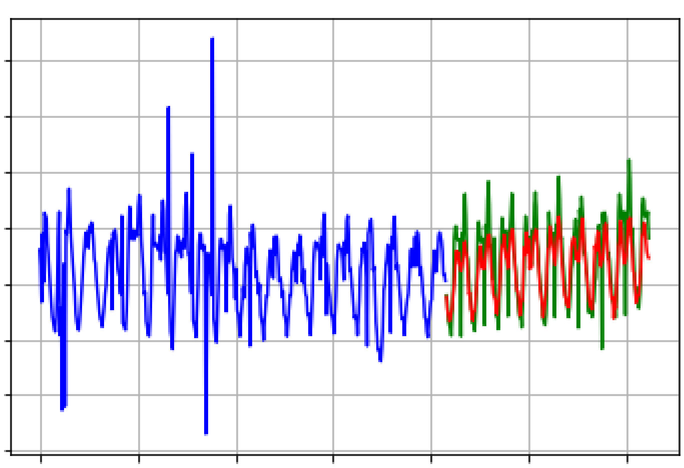
Dataset 21 mae : 2491.273 mse : 10805668.780 rmse : 3287.198 mape : 0.023



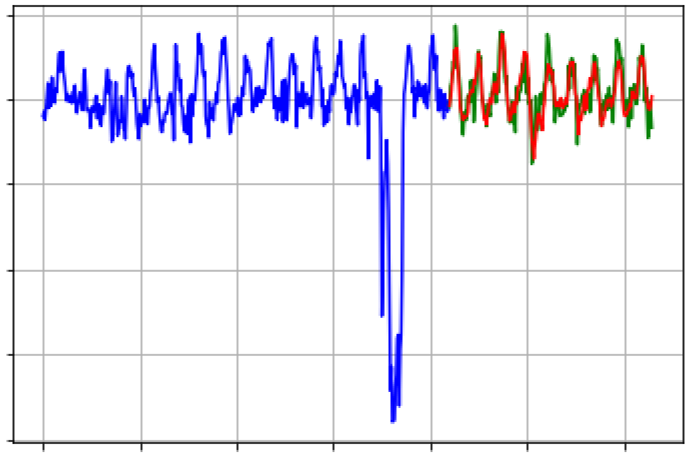
Dataset 22 mae : 2260.025 mse : 10328369.934 rmse : 3213.778 mape : 53.147



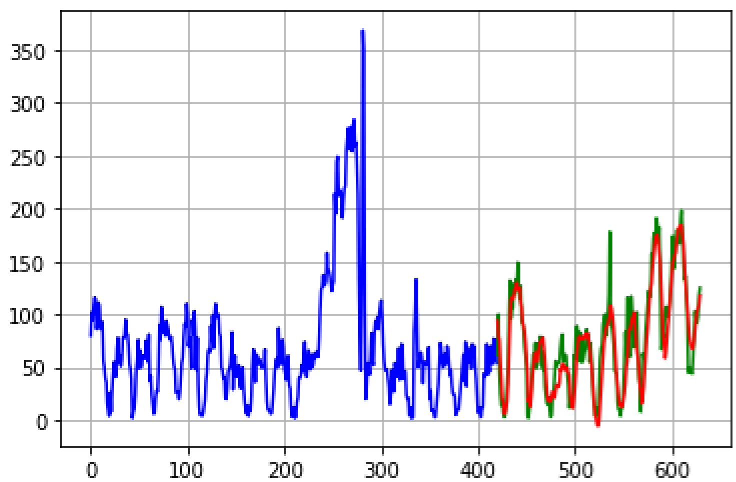
Dataset 23 mae : 3339.622 mse : 18960295.430 rmse : 4354.342 mape : 9.899



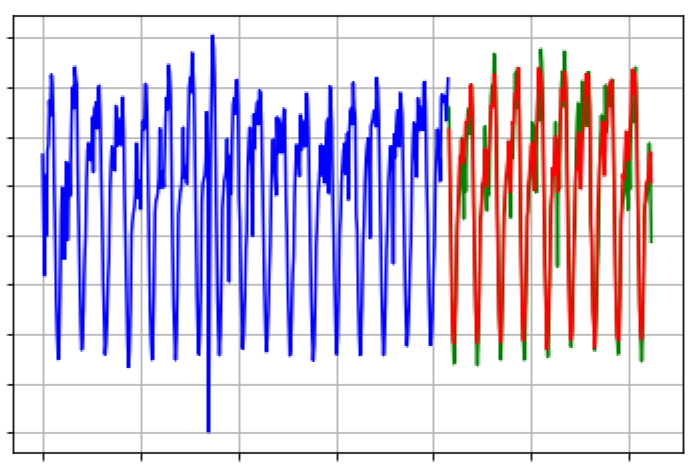
Dataset 24 mae : 0.285 mse : 0.125 rmse : 0.353 mape : 0.290



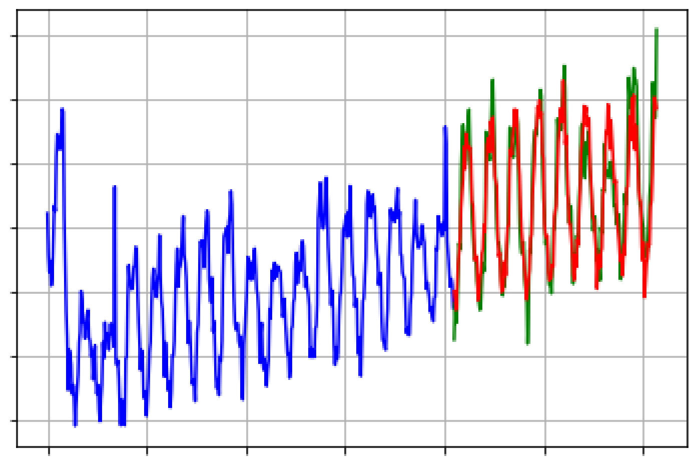
Dataset 25 mae : 15.557 mse : 392.762 rmse : 19.818 mape : 36.921



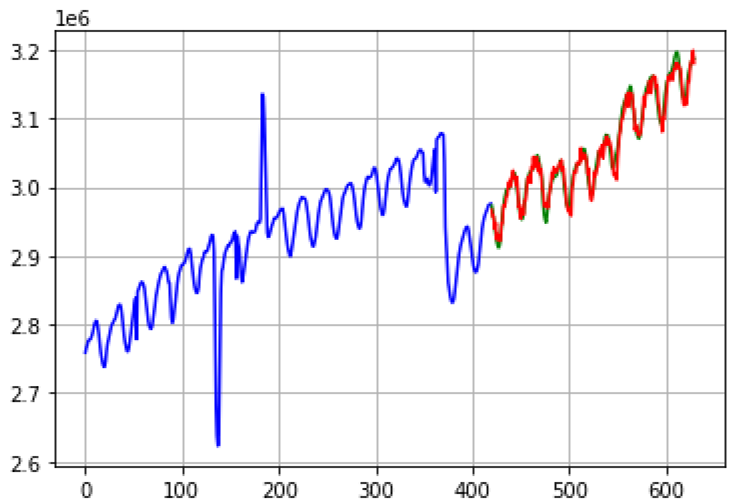
Dataset 26 mae : 921.955 mse : 1529339.654 rmse : 1236.665 mape : 8.372



Dataset 27 mae : 0.017 mse : 0.00049 rmse : 0.022 mape : 0.414

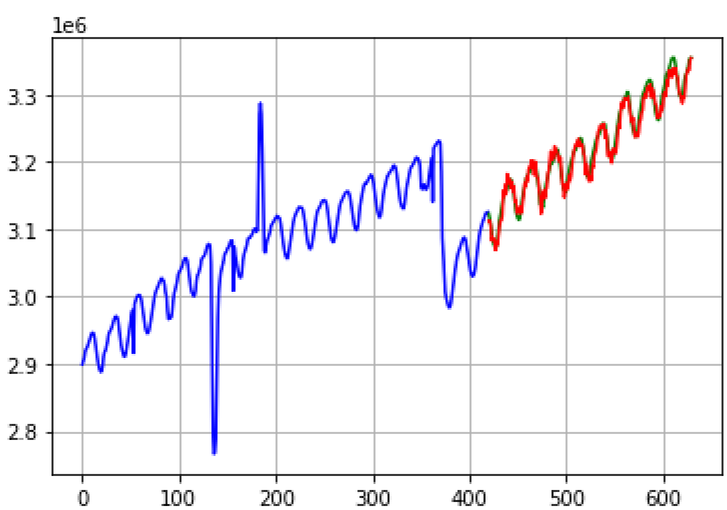


Dataset 28 mae : 8684.014 mse : 122237287.939 rmse : 11056.097 mape : 0.284

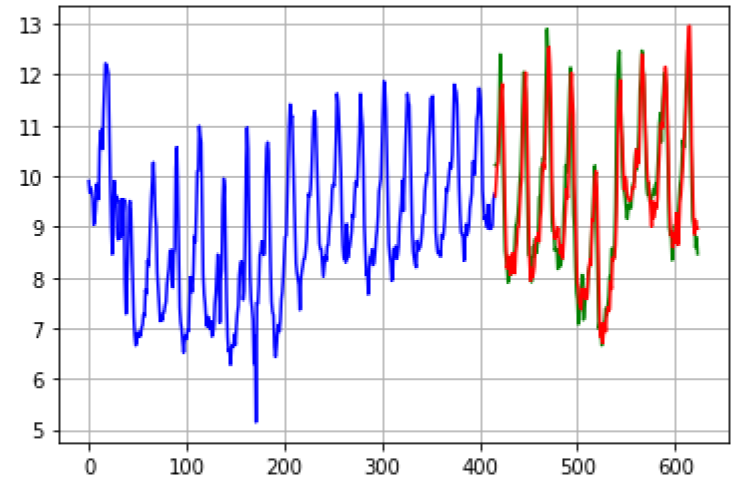


\

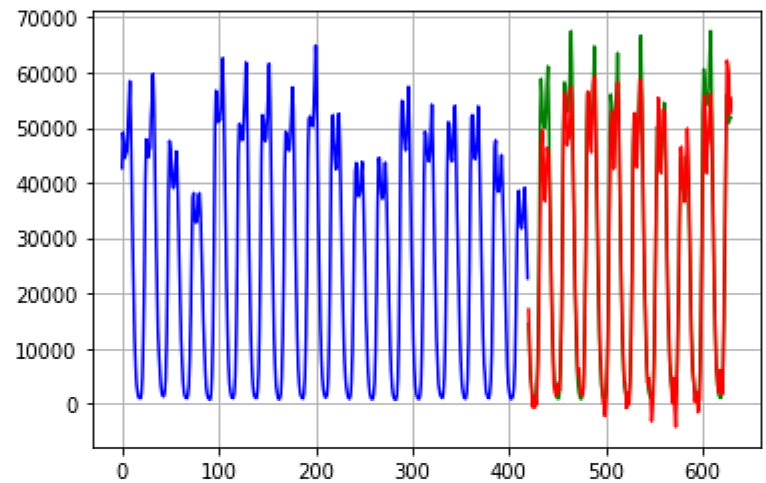
Dataset 29 mae : 8769.245 mse : 119233897.698 rmse : 10919.428 mape : 0.272



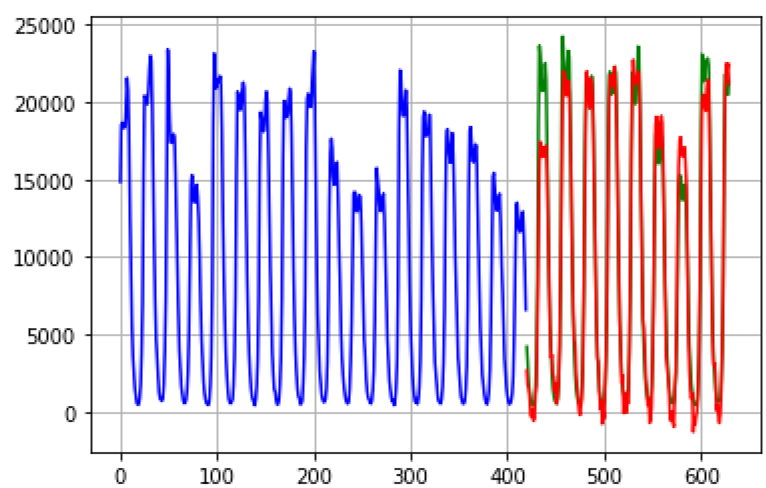
Dataset 30 mae : 0.412 mse : 0.306 rmse : 0.553 mape : 4.263



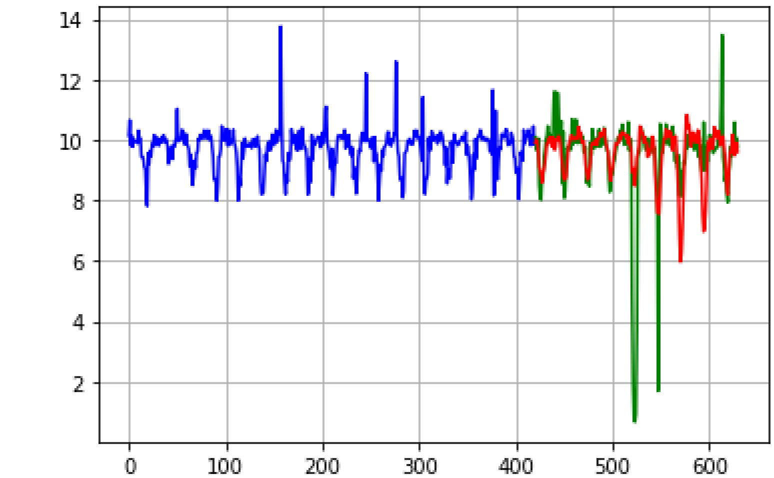
Dataset 31 mae : 2888.996 mse : 14826382.044 rmse : 3850.504 mape : 90.829



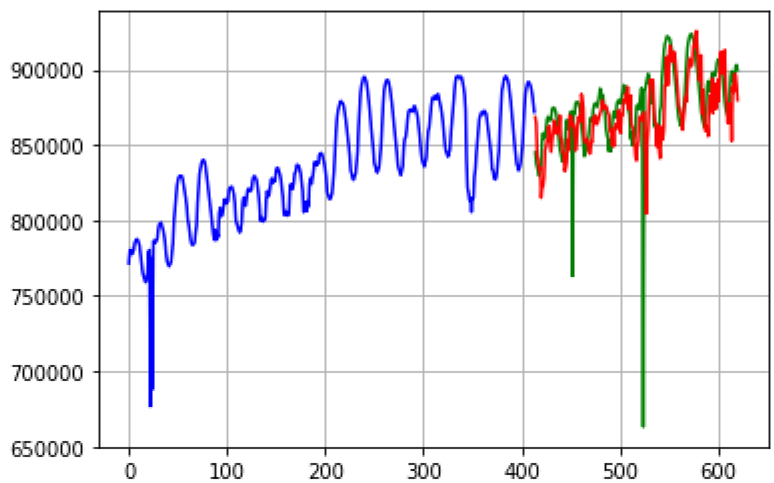
Dataset 32 mae : 1025.489 mse : 2232777.302 rmse : 1494.248 mape : 77.822



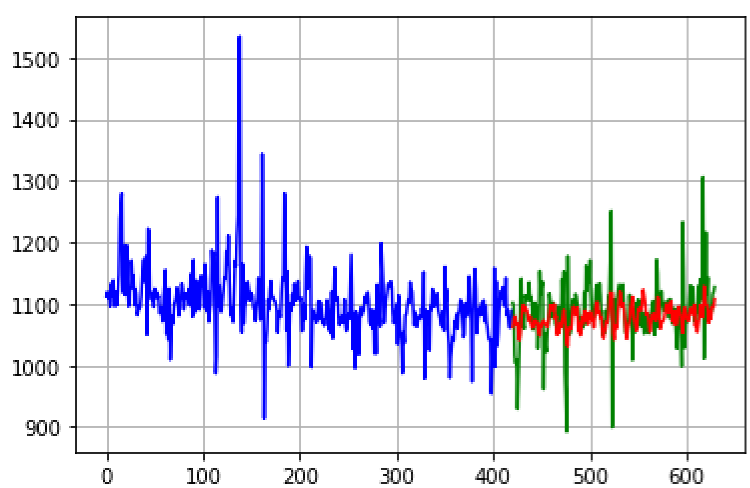
Dataset 33 mae : 0.711 mse : 1.974 rmse : 1.405 mape : 7.744



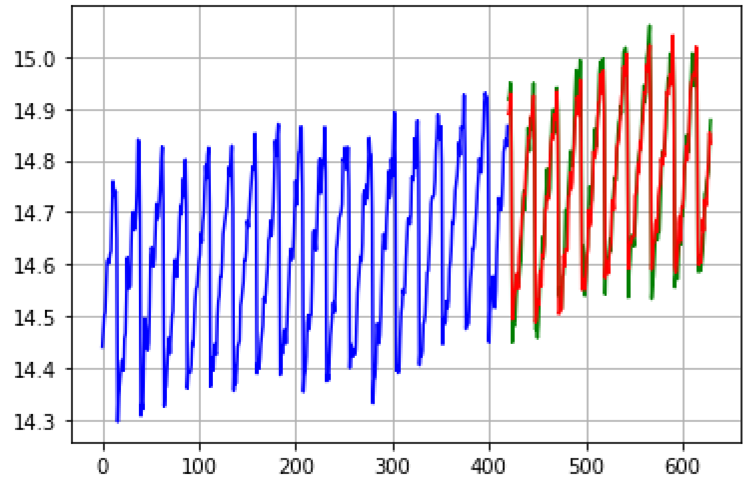
Dataset 34 mae : 15037.573 mse : 570036365.839 rmse : 23875.434 mape : 1.732



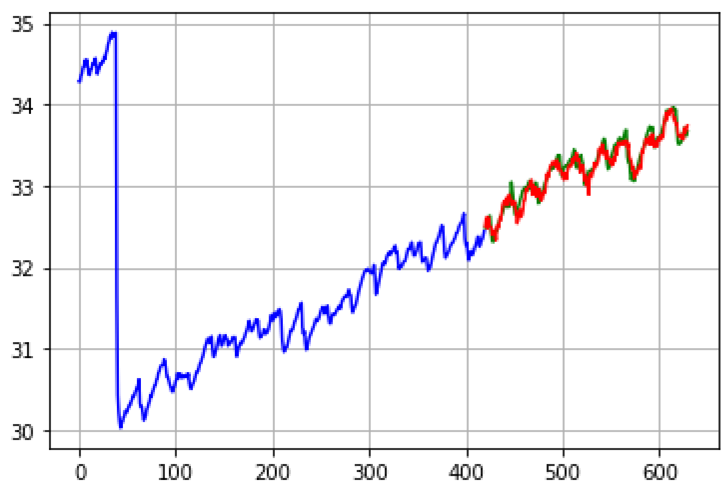
Dataset 35 mae : 31.233 mse : 2152.237 rmse : 46.392 mape : 2.890



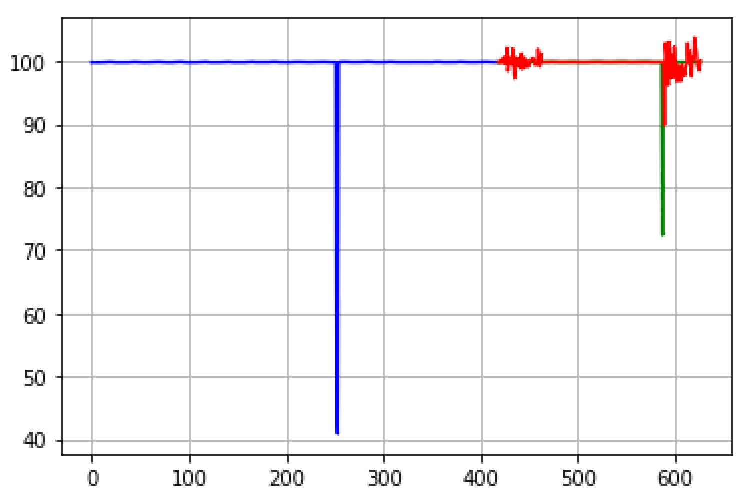
Dataset 36 mae : 0.027 mse : 0.001 rmse : 0.033 mape : 0.182



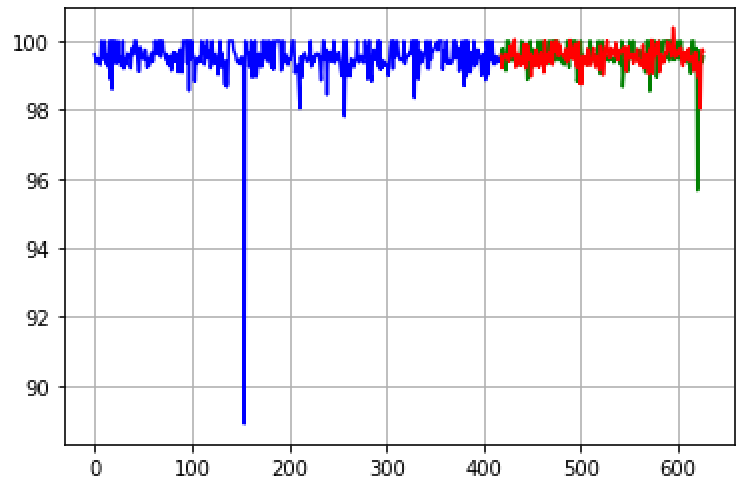
Dataset 37 mae : 0.077 mse : 0.010 rmse : 0.098 mape : 0.233



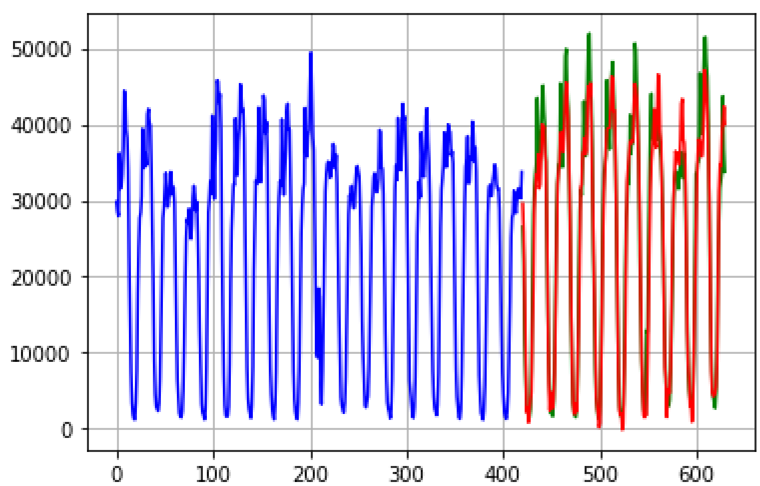
Dataset 38 mae : 0.547 mse : 4.134 rmse : 2.033 mape : 0.547



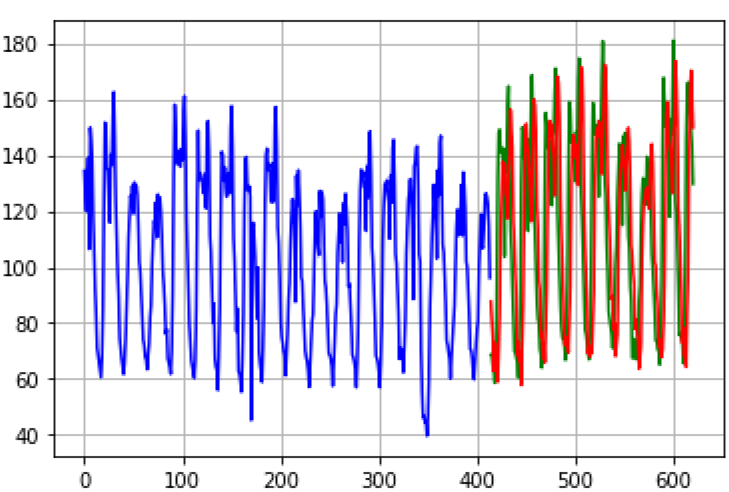
Dataset 39 mae : 0.269 mse : 0.173 rmse : 0.416 mape : 0.270



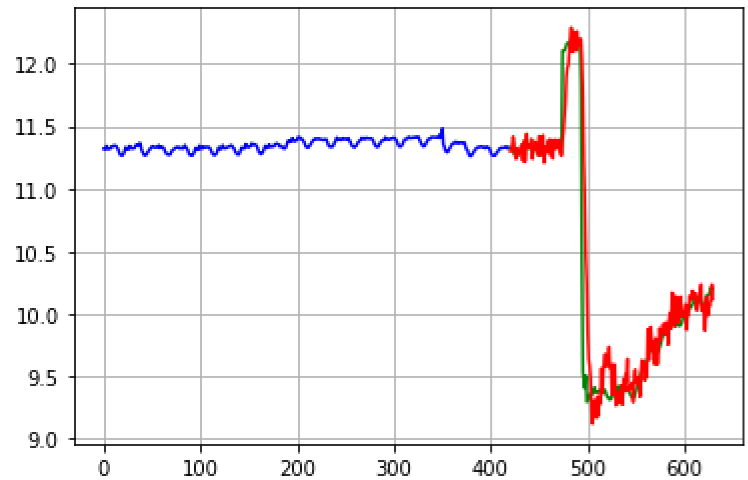
Dataset 40 mae : 2597.303 mse : 11918436.020 rmse : 3452.309 mape : 48.695



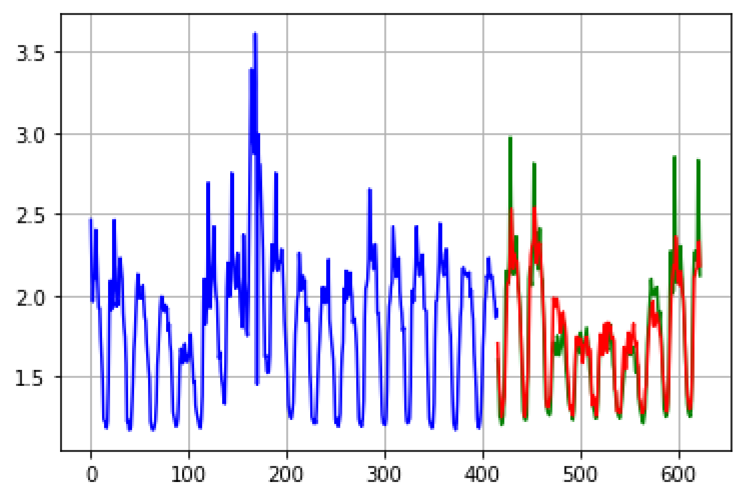
Dataset 41 mae : 24.339 mse : 980.742 rmse : 31.317 mape : 23.574



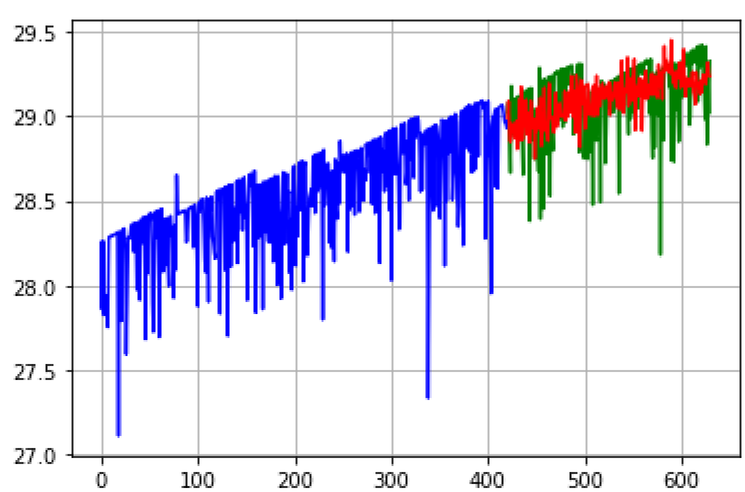
Dataset 42 mae : 0.141 mse : 0.092 rmse : 0.303 mape : 1.343



Dataset 43 mae : 0.103 mse : 0.022 rmse : 0.148 mape : 5.783

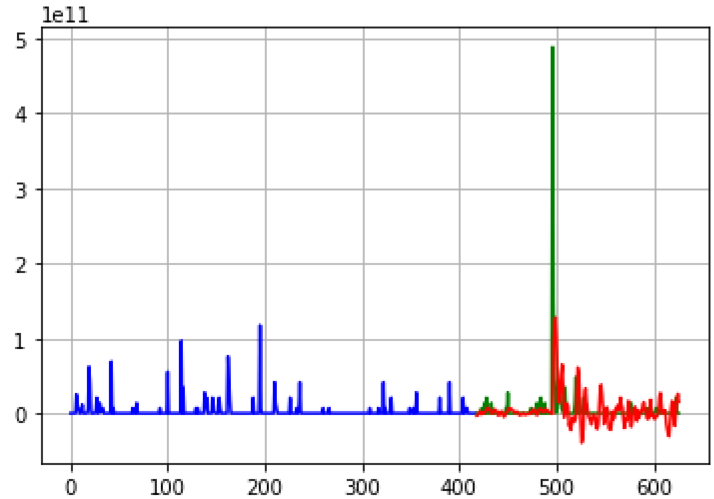


Dataset 44 mae : 0.172 mse : 0.046 rmse : 0.214 mape : 0.591

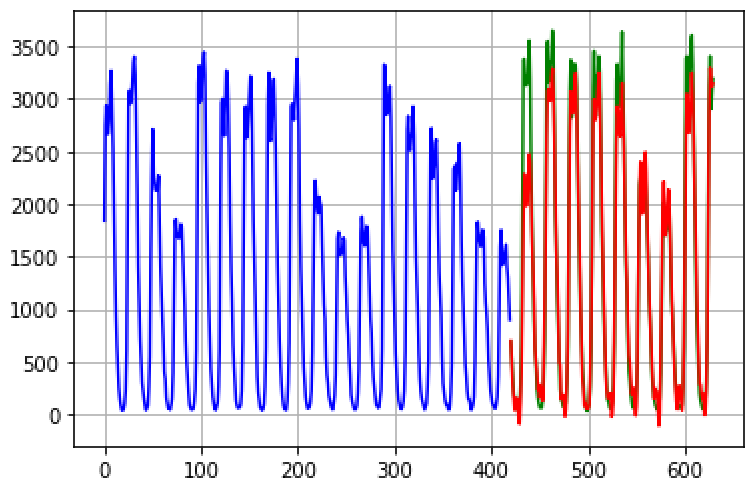


Dataset 45

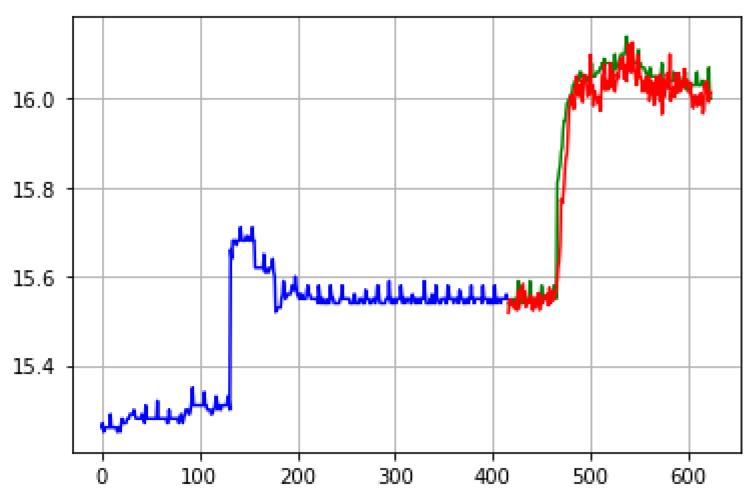
mae : 9758941503.267 mse : 1311904332163458269184.000 rmse : 36220219935.327 mape : 346.124



Dataset 46 mae : 197.128 mse : 93155.555 rmse : 305.214 mape : 139.969

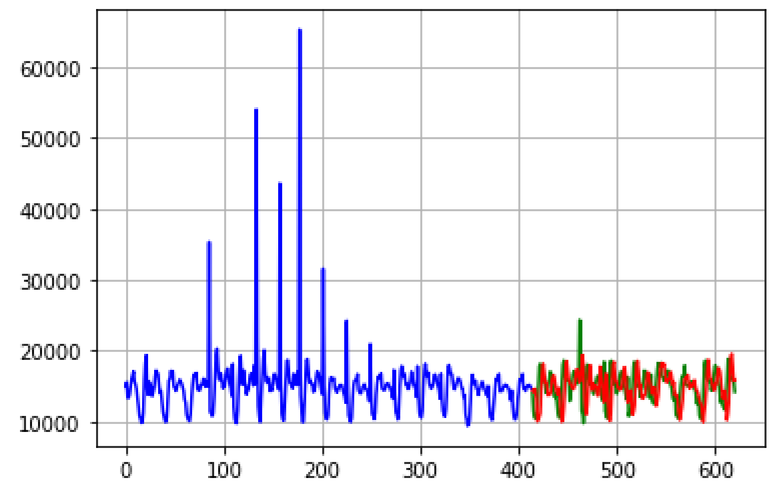


Dataset 47 mae : 0.034 mse : 0.002 rmse : 0.049 mape : 0.213



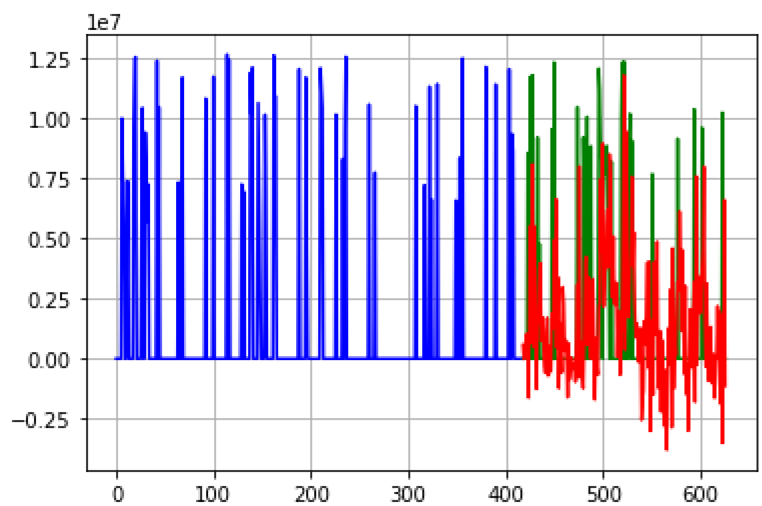
Dataset 48

mae : 2103.139 mse : 6996898.332 rmse : 2645.165 mape : 14.808



Dataset 49

mae : 2606481.150 mse : 13614490924644.990 rmse : 3689781.962 mape : 181.880



Thank You