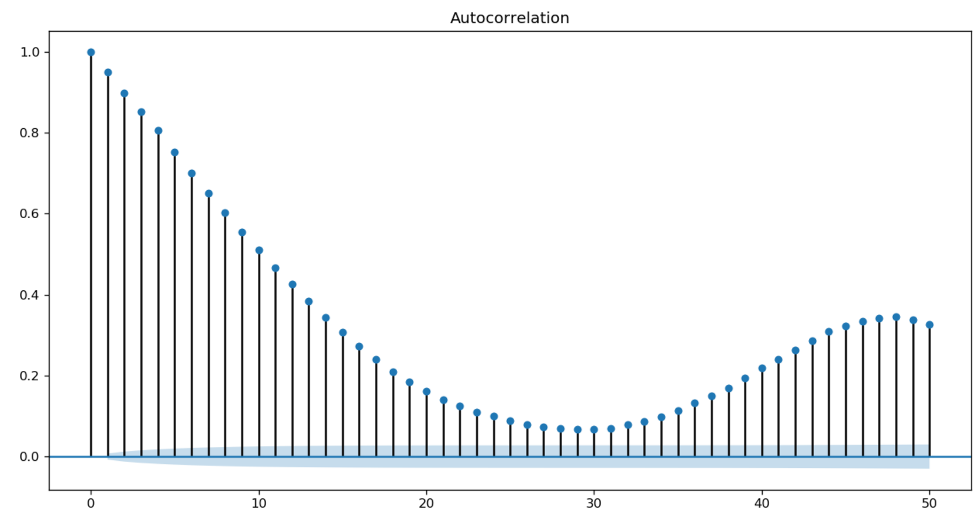
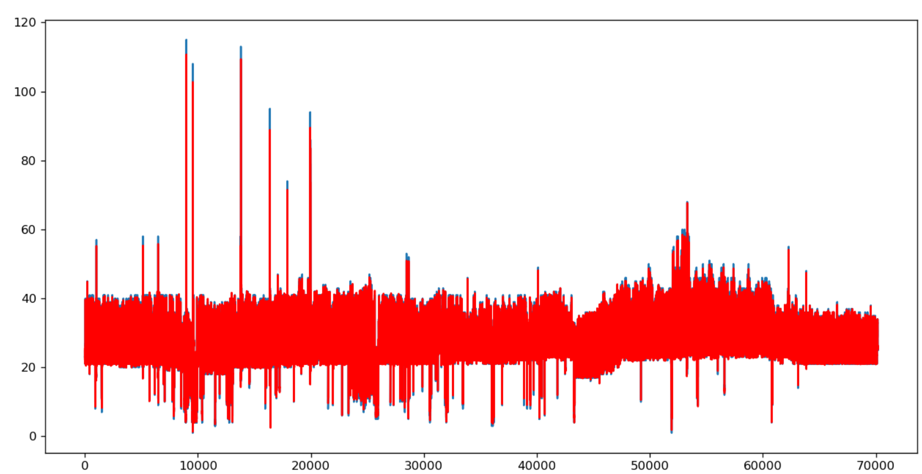
We have not run any feature selection algorithm because our data is time series data. We have an energy expenditure at 370 points. We chose one of these points and we made a prediction with the regression model. The point we selected is MT_002. We used autoregression to predict.

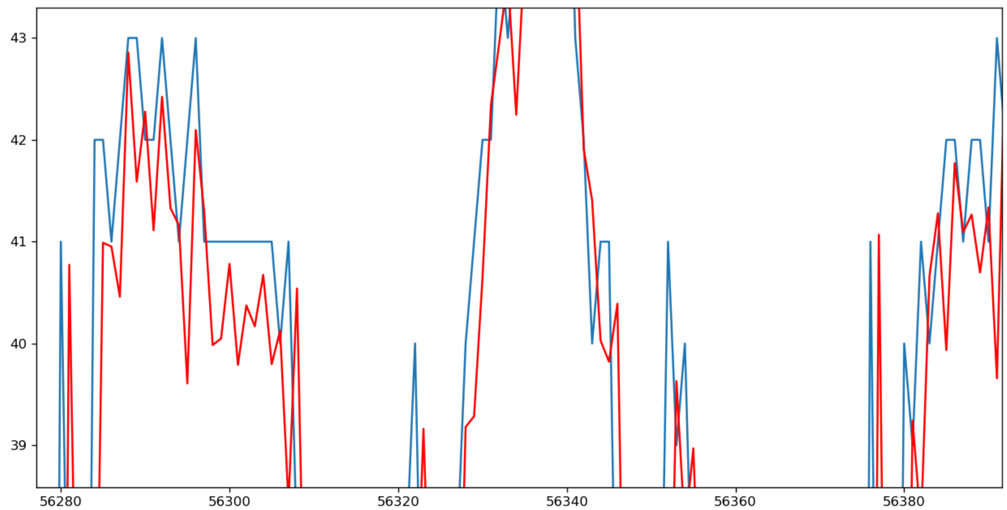
Auto Correlation Plot



Our Prediction Result



More Closer Look



Error Calculations

```
Test MSE: 4.402
Test MAE: 1.254
Test R2: 0.908
```

We had three years of data. We used the first two years for the train. We tried to predict the next year. We used an auto regression model for this.

A regression model, such as linear regression, models an output value based on a linear combination of input values.

For example:

yhat = b0 + b1*X1

yhat = b0 + b1*X1

Where yhat is the prediction, b0 and b1 are coefficients found by optimizing the model on training data, and X is an input value.

This technique can be used on time series where input variables are taken as observations at previous time steps, called lag variables.

For example, we can predict the value for the next time step (t+1) given the observations at the last two time steps (t-1 and t-2). As a regression model, this would look as follows:

X(t+1) = b0 + b1*X(t-1) + b2*X(t-2)

X(t+1) = b0 + b1*X(t-1) + b2*X(t-2)

An autoregression model is a linear regression model that uses lagged variables as input variables.

We could calculate the linear regression model manually using the LinearRegession class in scikit-learn and manually specify the lag input variables to use.

Alternately, the statsmodels library provides an autoregression model that automatically selects an appropriate lag value using statistical tests and trains a linear regression model. It is provided in the AR class.

We can use this model by first creating the model AR() and then calling fit() to train it on our dataset. This returns an ARResult object.

We use an alternative would be to use the learned coefficients and manually make predictions. This requires that the history of 29 prior observations be kept and that the coefficients be retrieved from the model and used in the regression equation to come up with new forecasts.

The coefficients are provided in an array with the intercept term followed by the coefficients for each lag variable starting at t-1 to t-n. We simply need to use them in the right order on the history of observations, as follows:

yhat = b0 + b1*X1 + b2*X2 ... bn*Xn

yhat = b0 + b1*X1 + b2*X2 ... bn*Xn