Does ML algorithm performance change over time?

|  |  |  |  |
| --- | --- | --- | --- |
| Doireann Hanley Trinity College Dublin Computer Science and French dhanley@tcd.ie | | Eunice OreOluwa Fasan Trinity College Dublin Computer Science and French fasane@tcd.ie | Sloane Reeves Trinity College Dublin Computer Science and German [sreeves@tcd.ie](mailto:sreeves@tcd.ie) |
|  |

# Introduction

# EXPERIMENTAL AND COMPUTATIONAL DETAILS

# Methodology

The dataset we used, titled ‘Monthly milk production: pounds per cow. Jan 62-Dec 75’ was sourced from the Time Series Data Library (TDSL)[1]. The dataset consists of 168 fact values in 1 time series. The time granularity is Month and the units used are pound per cow. The time range is Jan 1962 – Dec 1975.

Our data frame was created by reading the data into a .csv file. We parsed and formatted the dates accordingly and set the attribute ‘Month’ to be our index. We then simply plotted the data to look for potential patterns, evidence of trend and seasonality. We defined our problem and based on our time series data, proceeded to implement a range of quantitative forecasting methods. These included Moving Average, Exponential Smoothing, Double Exponential Smoothing, Holt-Winter’s model, ARIMA, Simple Lags/Linear regression, Ridge regression, Lasso regression and XGBoost.

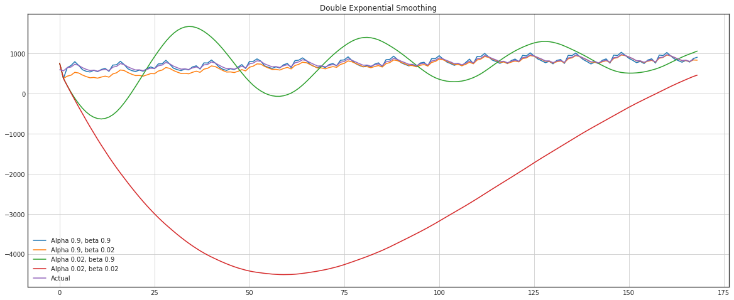
We chose to measure the quality of our forecasts using Mean Absolute Percentage Error (MAPE) as we consider it an interpretable metric that is particularly interesting because of its robustness to outliers.

We applied our baseline model, Moving Average, to smooth the original time series to identify trends. A close up of a map

Description generated with high confidence

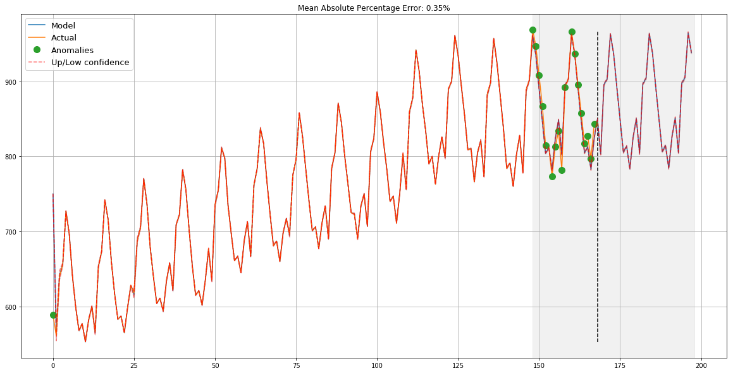
*Figure 1: Results from smoothing original time series with the previous 12 months and plotting the confidence intervals*

We then applied exponential smoothing and eventually extended this model to produce a double exponential smoothing whereby we tuned two parameters responsible for the series smoothing around itself, as well as the smoothing of the trend itself.



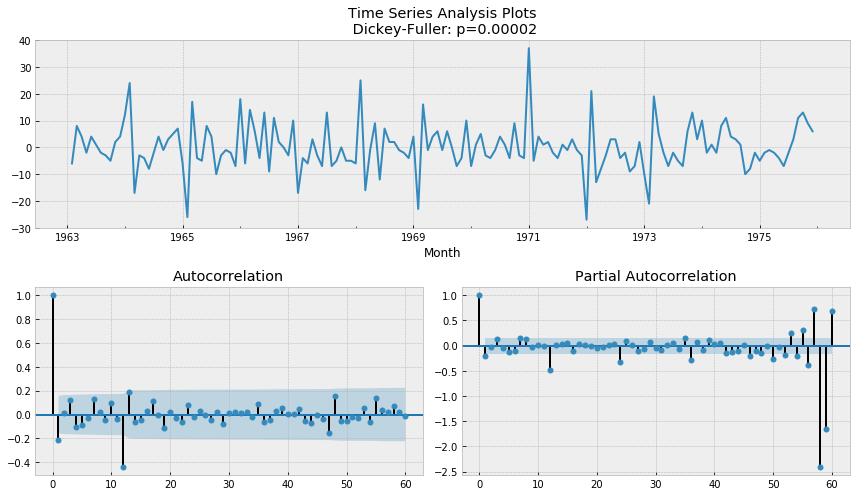
*Figure 2: Double exponential smoothing of data with alpha parameter values of 0.9 and 0.02 and beta values of 0.9 and 0.02*

The data was then split into training and test sets in the 80:20 ratio and thereafter we added a third component, seasonality, to our smoothing, effectively giving us a Holt-Winters model.



*Figure 3: Holt-Winter model approximating initial time series, capturing annual seasonality, overall upwards trend and even some anomalies with MAPE of 0.35%*

We then focused on achieving stationarity to increase our chances of predictions with minimal error. We used the Dickey-Fuller test, first order difference, ACF and PCF to reach stationarity.



*Figure 4: Stationary time series oscillating around zero. Dickey-Fuller test indicates stationarity as p< 0.05 i.e. p=0.00002 rejects null hypothesis that states the time series is non-stationary.*

We then trained a series of models and determined their MAPE error.

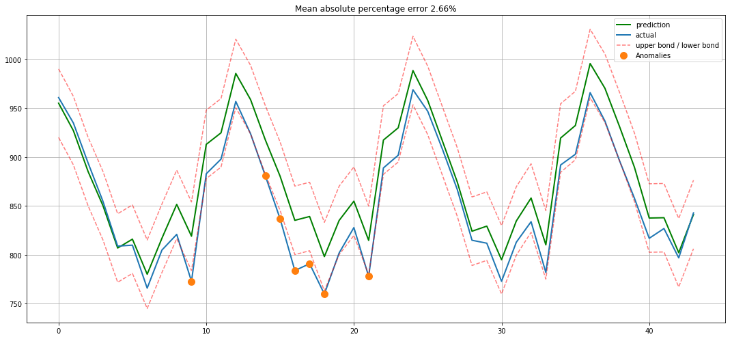
# RESULTS AND DISCUSSION

Holt-Winters model returned the least amount of error with a MAPE of 0.35% as opposed to Linear regression with simple lags which returned a MAPE of 2.43% and other algorithms such as Lasso and Ridge regression as well as XGBoost which returned errors of 2.66%, 2.45% and 3.45% respectively.

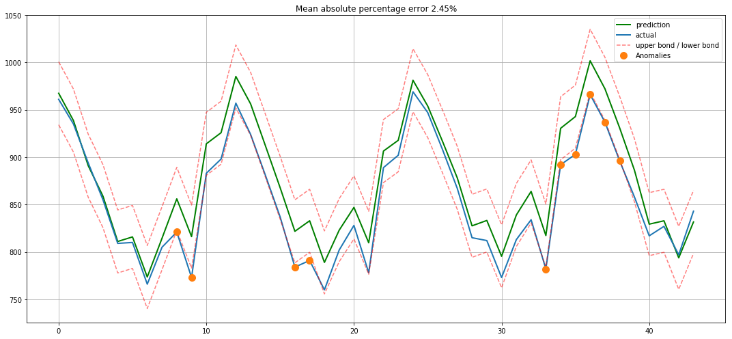
A close up of a map

Description generated with very high confidence

*Figure 5: Simple lags and linear regression plot giving approximated predictions with an error of 2.43%*



*Figure 6: Lasso regression prediction with MAPE of 2.66%*



*Figure 7: Ridge regression prediction with MAPE of 2.45%*

A close up of a map

Description generated with high confidence

*Figure 8: XGBoost model prediction with MAPE of 3.45%*

If we were to rate these models based solely on their individual MAPE scores, Holt-Winters would be a clear winner. We must however take into account that these error values are only an average of the cumulative monthly errors. If we were to base our overall winner on plotted predicted values, expected values and most importantly error per time interval i.e. each month of the year, it’s possible that the Holt-Winters model could be dethroned as there could be a decreasing trend with its predictions and an increasing trend with another model’s prediction.

# Limitations and outlook

One crucial step to help us better evaluate our chosen models and their performance is to plot the error per time interval i.e. each month, as opposed to an overall single number metric telling only of the average error over the previous 12 months. Errors in our code denied us this insight therefore we will aim correct our code and produce the desire result as part of our second submission.

# CONCLUSIONS

A HEADINGS IN APPENDICES

A.1 Introduction

A.2 Experimental and Computational Details

A.3 Results and Discussion

A.4 Conclusions

A.5 References