ML1819 Research Assignment 1

Team 26, Task 103 - Does ML algorithm performance change over time?

Eunice OreOluwa Fasan – 14311471 Methodology, Limitations and Outlook, Results and Discussion
Doireann Hanley – 15317527 Introduction, Limitations and Outlook, TITLE
Sloane Reeves – 15320345 – Related Works, Limitations and Outlook

Word Count - 1404

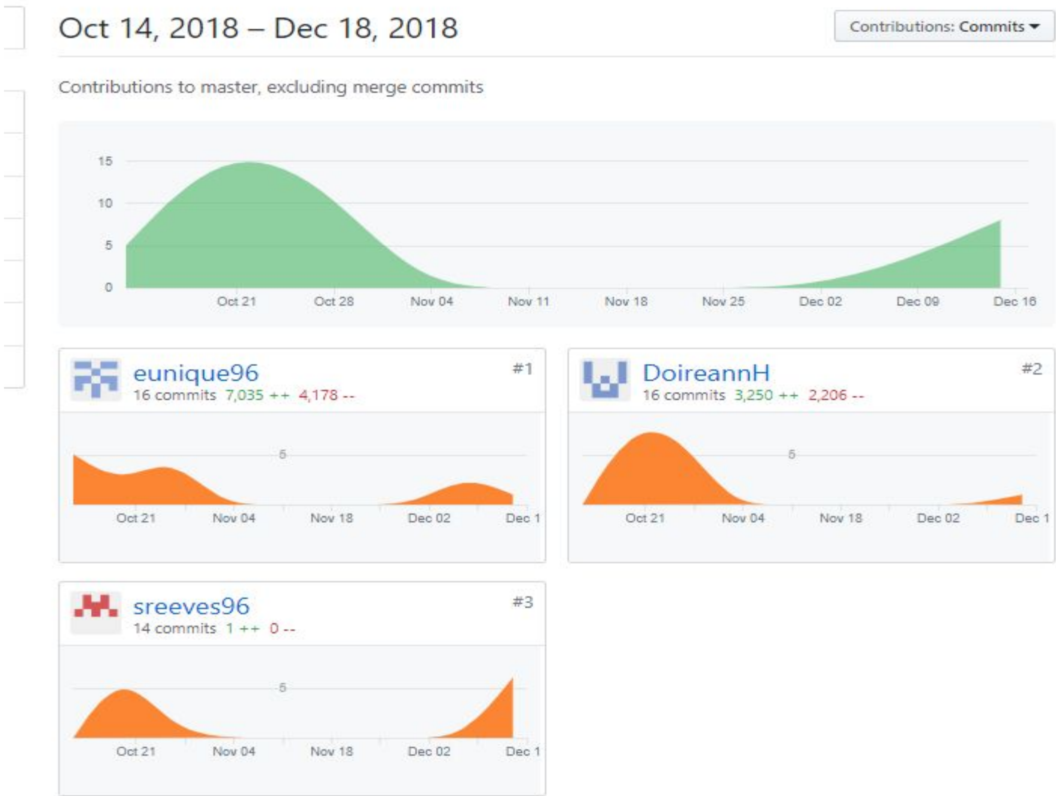
Source Code:

https://github.com/eunique96/ML1819--task-103--team-26.git

https://github.com/eunique96/ML1819--task-103--team-26/blob/master/ML1819--task-103-team-26_II.ipynb

Activity:

https://github.com/eunique96/ML1819--task-103--team-26/graphs/contributors



We would like to note our overall contribution is lower on github for this second part of the assignment as most of our work was continued through google drive.

Single and ready to mingle?

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

1 INTRODUCTION

Time can have a major influence on a model's predictive performance. If a model is tested on a new, rapidly evolving dataset, its performance may drastically deteriorate over time. 'Model degradation' is an issue which is evident in information security today and is essential in understanding a model's predictive performance. Often, researchers use metrics such as mean absolute error (MAE) or root mean square error (RMSE) to portray how well an algorithm has performed on average over the period of data collection. The question of whether single-number metric results are a true reflection of an algorithm's performance, as it pertains to time, is one that has only gained traction in recent years. In an attempt to answer this question we calculated the RMSE at each time interval for our chosen time series. We implemented four different methods - Naive Approach, Moving Average, Simple Exponential Smoothing and Holt's Linear Trend and provided graphs to illustrate our findings. In this paper, in addition to presenting overall RMSE scores for each algorithm, we also hope to show how each algorithm's RMSE develops over a period of time with respect to the dataset and as a result draw a more meaningful conclusion as to which algorithm may best suit the dataset.

2 RELATED WORKS

While we aim to see how Machine Learning algorithms are affected over time, traditionally time has been only considered in seeking to predict the preference a user would have to an item, i.e. recommender-system research. Time is considered important in this case in marking the changes in user preference over time. Lee et al (2008) for example use a time based approach to effective recommender systems using implicit feedback [4]. Their focus lies on adjusting algorithms to make them consider time and improve the effectiveness of the algorithms in accordance with the users wishes [3]. An example of a temporal feature as such is seasonality. Seasonality is often considered as an important contextual feature, i.e it is something to be considered that one should use a different algorithm in winter than in summer. However, it seems from our study that all of the research concerning time attempts to use time to improve machine learning algorithms as opposed to a comparison study for machine learning models for regression, which we aimed to conduct. To the best of our knowledge, most of the works that relate to our studies are focused on the effectiveness of a particular machine learning algorithm over a predictive time-series. Kyoung-jae Kim et al (2002) tries to evaluate the feasibility of using Support Vector Machines in order to predict a financial time series [2]. This study examines the feasibility of applying SVM in financial forecasting by comparing it with back-propagation neural networks and case-based reasoning. Another study conducts time-series forecasting focusing on using Simple Exponential Smoothing, one of the models we observe in our study [5]. The work we found most closely related to ours, Atiya and Ahmed et al(2010), utilises a comparable method to ours aiming to answer a similar question, as they present a large scale comparison study for the major machine learning models for time series forecasting. The models that they use differ from ours using 'multilayer perceptron, Bayesian neural networks, radial basis functions, generalized regression neural networks, K-nearest neighbor regression, CART regression trees, support vector regression, and Gaussian processes'. They also apply the models to monthly time series data however on a larger scale, with 'around a thousand time series'[1]'.

3 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) [6]. We created train and test files for modeling and used the first 12 years of data (Jan.1962 - Dec.1974) as the training set and the last year of data (Jan. 1975 - Dec.1975) as the test set. The data was framed, parsed and formatted accordingly and we proceeded to implement a range of forecasting methods. The Root Mean Square Error (RMSE) metric was used as our performance measure.

We began by establishing a baseline of performance which was the Naive Approach. In this method, the observation from the previous time step was used as the prediction for the observation at the next time step. The results of this can be seen in Figure 1.

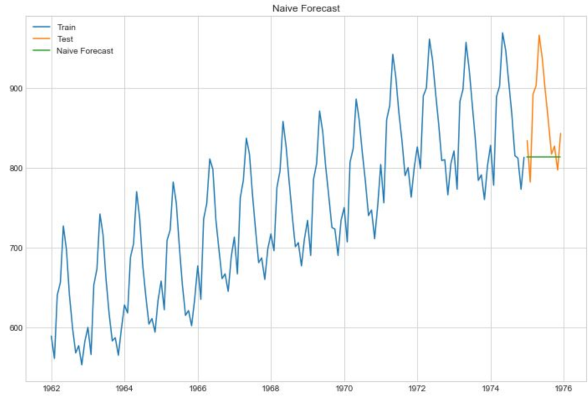


Figure 1: Results from plotting training and test data using Naive approach model as predictor.

The Naive approach model achieved an overall RMSE of 48.389 meaning that on average the model was wrong by about 48 pounds of milk produced for each prediction made. Following that we used the Moving Average model, the results of which can be seen in Figure 2. It predicted the next values in the time series based upon the average of a fixed number of previous values.

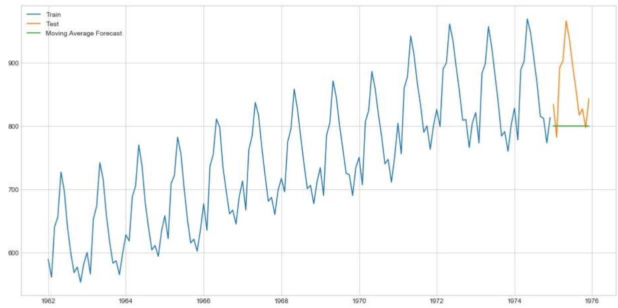


Figure 2: Results from plotting training and test data using Moving Average model as predictor with fixed number of previous values equal to 3.

Moving Average achieved an overall RMSE of 99.830.
Next we considered the Simple Exponential Smoothing (SES) method. This technique calculated the next time step as an exponentially weighted linear function of observations at prior time steps. A smoothing parameter (alpha) was used and the results can be seen in Figure 3 below.

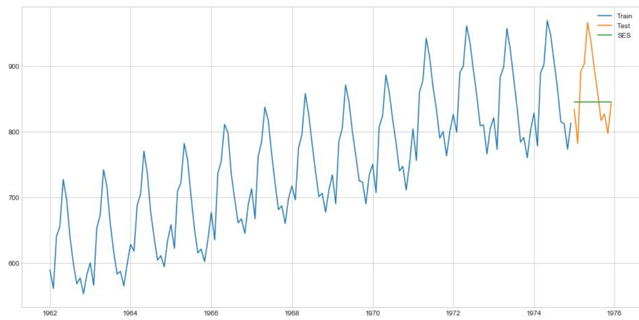


Figure 3: Results from plotting training and test data using Simple Exponential Smoothing as predictor with alpha equal to 0.08.

The overall RMSE achieved by the SES method was 56.979. Up to this point we can see that based upon the overall RMSE values, the Naive method outperformed both the Moving Average Method and the SES method for this particular dataset.
Lastly we chose to implement the Holt's Linear method. This technique simply extended the SES method to allow the forecasting of data with a trend. The results of this method applied to the dataset are shown in Figure 4.

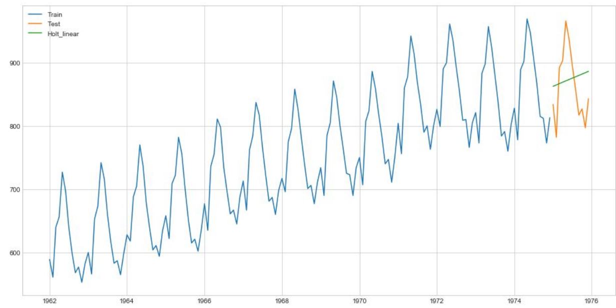
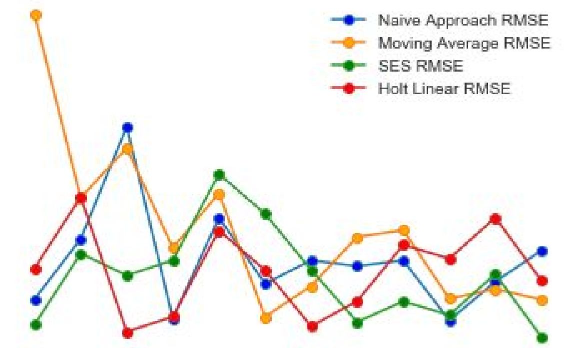


Figure 4: Results from plotting training and test data using Holt's Linear Trend method as predictor.

The Holt's Linear method achieved an overall RMSE of 57.516.

4 RESULTS AND DISCUSSION

If we were to rate these models based solely on their overall individual RMSE scores, we would be inclined to choose the Naive approach as winner since it achieved the lowest score. However we took into account that these scores were only an average of the cumulative monthly errors. Our argument therefore was that presenting overall RMSE was not truly indicative of the algorithms' performances over the 12 months used to test the models. To better gauge the reliability of our results, especially as it pertained to deciding on the best model for the dataset, we plotted the RMSE values for all 12 months of the test data for each model. The results of this plot can be seen in Figure 5.



	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12
Naive	21.0	52.0	110.0	11.0	63.0	29.0	41.0	35.0	41.0	10.0	30.0	46.0
MA	237.3	103.7	140.0	67.0	107.0	16.7	39.3	75.0	50.0	30.0	37.0	29.3
SES	7.5	44.5	33.2	41.0	55.6	65.1	36.1	9.2	19.5	12.7	33.9	1.4
Holt	36.9	73.7	4.1	11.9	56.4	35.9	6.9	19.9	45.9	41.9	63.1	30.5

Figure 5: RMSE of Naive approach, Moving Average, SES and Holt's Linear models over time. m1 to m12 represent the months based on the test data. Values in cells represent the RMSE scores of each month for all models implemented.

We can infer from Figure 5 and previous figures that although the Naive approach achieved the lowest overall RMSE score, it had the potential for degradation. Over time it appeared that the SES model ended with the lowest RMSE and from looking at the trend, it's probable that the SES model's decrease in RMSE would continue in the future and may be the model best suited for the dataset.

It is evident that single metrics are only a partial reflection of an algorithm's performance over time and in order to make a more informed decision on performance, at least when it concerns time series, it would make sense to calculate and plot metrics for each time interval to obtain a more complete picture of the algorithm's performance and tendencies over time.

5 LIMITATIONS AND OUTLOOK

The dataset which we chose to train our models on was a concern for us throughout this project. Given that we are analysing different methods over a period of time, if we were to continue our research we would train our models on a dataset with a greater number of fact values. The 'Monthly Milk Production' dataset which we chose to use only has 168 fact values, over a period of 13 years. Ideally, the next step in improving our research would be to implement our models on a dataset with at least 1000 unique values.

REFERENCES

[1] Ahmed, N., Atiya, A., Gayar, N. and El-Shishiny, H. (2010). An Empirical Comparison of Machine Learning Models for Time Series Forecasting. Econometric Reviews, 29(5-6), pp.594-621.

[2] Kim, K. (2003). Financial time series forecasting using support vector machines. Neurocomputing, 55(1-2), pp.307- 319.

[3] Koren, Y. (2010). Collaborative filtering with temporal dynamics. Communications of the ACM, 53(4), p.89.

[4] Lee, T., Park, Y. and Park, Y. (2008). A time-based approach to effective recommender systems using implicit feedback. Expert Systems with Applications, 34(4), pp.3055-3062.

[5] Ostertagová, E. and Ostertag, O. (2012). Forecasting using simple exponential smoothing method. Acta Electrotechnica et Informatica, 12(3).

[6] Time Series Data Library, Agriculture, Source: Cryer (1986), in file: data/milk, Description: Monthly milk production: pounds per cow. Jan 62 - Dec 75