

COMPARISON OF TIME SERIES PREDICTION OF HEALTHCARE EMERGENCY DEPARTMENT INDICATORS WITH ARIMA AND PROPHET

Diego Duarte¹ and Julio Faerman²

¹FLAS (Faculty of Liberal Arts & Sciences) University of Greenwich,
London, UK

²Universidade Autònoma de Barcelona, Plaça Cívica, Barcelona, Spain

ABSTRACT

Predicting emergency department (ED) indicators in time series may benefit hospital planning, improving quality of care and optimising resources. It motivates analysis of models that can forecast relevant KPIs (Key Performance Indicators) for identifying future pressure. This paper analyses the Autoregressive Integrated Moving Average (ARIMA) method in comparison to the analysis of Prophet, an autoregressive forecasting model based on Re-current Neural Networks. The dataset analysed is formed by hourly valued hospital indicators, composed by Wait to be Seen Major in ED, Number of Attendances Major in ED, Unallocated Patients in ED with a DTA and Number of Beds Available on Medical Acute Unit. A comparison of predictions models ARIMA and Prophet is the focus. Each model is designed to provide better predictions for different time series characteristics. Measurements of best prediction for each indicator are based in accuracy, reliability bands and indicator meta information.

KEYWORDS

Predicting, Healthcare, ARIMA, Prophet & Time series

1. INTRODUCTION

In the United Kingdom, due to factors such as population, economy, weather, and others, the NHS (National Healthcare System) hospitals often experience high pressure in the ED (emergency department), having to compromise the very high standards set for the country in this area. Predicting ED acute units (hospitals) indicators in time series help understand pressure levels in advance and may improve quality by planning ahead, optimising staff shifts and setting escalation plan actions.

When analysing health care system, increasing importance has been focused to predicting patient arrivals related data in acute units, especially the prediction of ED patient attendances and waiting times. According to ED data, patients arrive in an unplanned manner, often beyond control of the hospital, and shows strong periodical, seasonal and stochastic fluctuations driven by factors such as climate, epidemic disease, the type of the day and socio demographic effects [1]. Based on these features, researchers used different methods to predict the number of ED visits hourly, daily and monthly [2], [3].

Conversely predicting KPIs in ED depends on several variables [4], [5], typically, ED deals with life threatening emergencies of all kinds, e.g.: car accidents, weather related disease, average stress rate, pollution, working hazards, aging, growing and migrating population, and countless

other factors. Multivariate models tend to consider only a number of measurable of dependent and independent variables out of countless, in ED that is a dangerous assumption, once series of non-captured factors may drive multi-variate predictions to the wrong side, such as inconsiderate variables having higher correlation levels, variable correlation lags, noise of each considerate variable, and others [6].

For this type of prediction, a common strategy is to auto correlate data. All countless variables have been considerate in forming past values, thus predicting models yield such data as baseline, to calculated future values in time series. Apart from the base-line, other variables, dependent or not represented by other indicators may be added in the calculations depending on the level of covariance found between them.

Accurate and reliable forecasts of the different types of KPIs can help to allocate the key healthcare resources scientifically like nurses, doctors, medical equipment at moment that is most needed [2], as well as not allocating them in non-critical periods. This justifies investment to make precise predictions of ED data, in order to provide hospital managers with the right information to make the best decisions to deliver the expected healthcare demand effectively and timely [1].

It is beneficial for healthcare managers to gain knowledge of short-term variations in ED. As a linear model, ARIMA fits the purpose to capture linear patterns in a time series efficiently, so most studies adopt it to describe the relationship between the variables or use them as the benchmark to test the effectiveness of combined models with mixed results [7].

Sun et al. [8] and Milner et al. [9] developed ARIMA models for forecasting patients' attendances in ED of hospitals to verify the fitness of the model in the context, readily available tool for predicting ED workload. Milner et al. [9] used ARIMA to fore-cast yearly ED patient attendances in Trent region in UK in a having found that fore-cast for attendances was reasonably close to the actual figure, overestimating it by 3 per cent. Sun et al. [3] found improvements in ARIMA model for daily patient attendances in Singapore General Hospital by combining it with other variables like air quality, public holidays and periodicities.

Prophet is a forecasting model developed by Facebook Research, an R&D branch of the company motivated by finding innovative solutions. Taylor et al. [10] observed ARIMA forecasts are prone to large trend errors when there is a change in trend near the cut-off period and fail to capture any seasonality. Experimenting with Prophet Taylor et al. [10] predicted daily number of events on Facebook with 25% more accuracy than ARIMA for predictions in a 30 days horizon.

Facebook daily number of events is an indicator with similar time series characteristics in trend, seasonality, and holidays, that are base of Prophet model decom-position, what motivated this study to apply the model to healthcare data.

2. RELATED WORK

Forecasting indicators in time series is a problem in which researchers proposed different models and solutions. Each model benefits different situations and types of time series data. Recently the use of RNN (Recurrent Neural Network) model likewise Prophet, has gained momentum by publications evidencing enhanced results compared to traditional statistical methods. Yi, Bu and Kim (2019), proposed a model that has a trend part, considering slope, duration and adjustment, in which a simple Recurrent Neural Network (RNN) is applied to learn them, complimentary, it was considered a non-trend part in the time series, a mix of trend and non-trend parts

demonstrated a better performance in a selection of real-life examples in comparison to other averaging-based statistical forecast methods and recurrent methods [11].

Such improvements are not present in other circumstances using RNN, Elliot and Hsu (2017) used a Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) for predicting financial time series of stock prices, proposing an architecture of an input layer linked to a layer of LSTM nodes, connected to a 1 node dense layer as output, in their study LSTM-RNN fail to better the forecast accuracy of 1 dimensional time series in comparison to traditional linear models [12].

3. ARIMA (AUTOREGRESSIVE INTEGRATED MOVING AVERAGE)

A basic requirement for predicting time series with ARIMA is the time series to be stationary or at least trend stationary [3], [7]. A stationary series has no trend, and variations around the mean have a constant amplitude, and it varies in a pattern [1], [13].

ARIMA model depends on order parameters (p, d, q), these parameters values are defined by calculations and analysis, process of fitting an ARIMA model is widely known as the Box-Jenkins method. Parameters p, d, and q are non-negative integers, p is the order (number of time lags) of the autoregressive model, d is the degree of differencing (number of times the data have had past values subtracted if random walk is chosen), and q is the order of the moving-average model.

Autoregressive Integrated Moving Average ARIMA (p, d, q) is the integration of AR (Auto Regressive) and MA (Moving Average), with p autoregressive terms and q moving average terms [12].

$$\hat{y} = c + \epsilon_t + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i}$$

Error terms ϵ_t are generally assumed to be independent, and homogeneously distributed random variables, c represents a constant variable, y are actual values in the temporal moment t.

Each of time series in analysis, has been through the ARIMA / Box Jenkins model in the three-stage iterative modelling approach (see Figure 1):

1. Model identification, making sure variables are stationary, if not, apply differencing method such as random walk, identifying seasonality, plot ACF (Auto Correlation Function) and PACF (Partial Autocorrelation Function) to decide if stationary and the ARIMA parameters p, d, q [14].
2. Parameter estimation, realize predictions (future and historical) generally using methods such as Maximum Likelihood Estimation or Non-linear Least Squares Estimation and identified parameters [14].
3. Model checking, testing if predictions adhere to specifications of stationary univariate process, in particular, residuals have to be within acceptable range and must be independent of each other, mean and variance should be constant over the time. If the model is inadequate, return to stage 1 and apply a different method do differencing in an iterative manner [14].

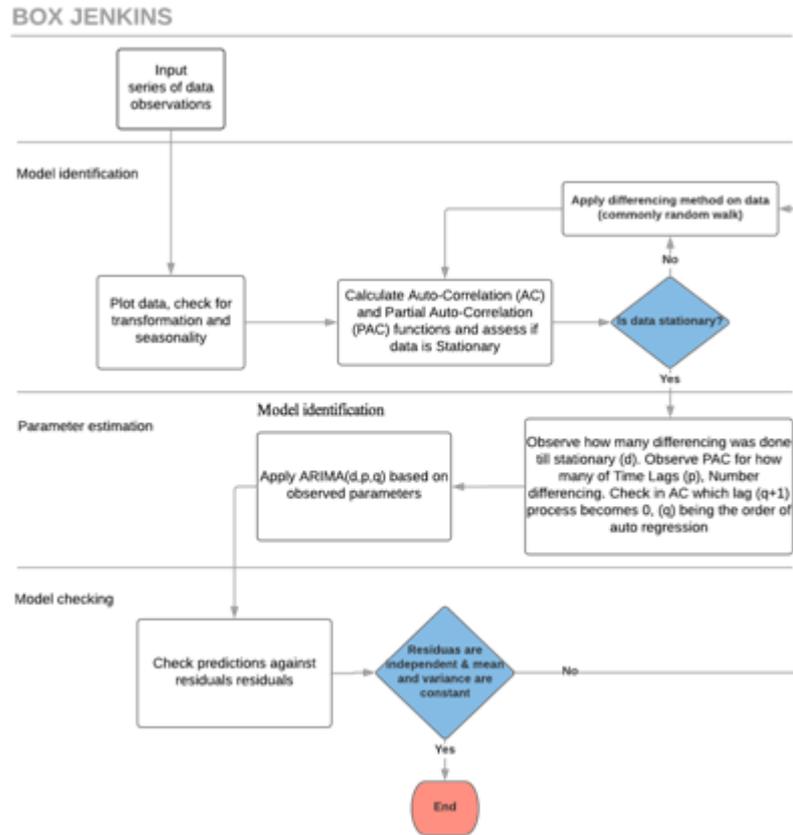


Figure 1. Box Jenkins iterative fitting model

4. PROPHET

Prophet is a Facebook Research forecasting model, according to Facebook, forecasting is a data science task that is central to many activities within a large organization such themselves, they consider a must to engage in capacity planning and efficiently allocate scarce resources and goal setting in order to measure performance relative to a baseline. Delivering high quality forecasts is a difficult task for either machines or for most analysts [10].

When fitting MFT ED KPIs in Prophet is driven by the idea not every prediction problem may be addressed by the same solution. It has been created with purpose of optimizing business forecast tasks encountered at Facebook, that usually have the characteristics:

- Hourly, daily, or weekly observations with at least a few months (preferably a year) of history.
- Strong multiple “human-scale” seasonality: day of week and time of year.
- Important holidays that occur at irregular intervals that are known in advance (e.g. the Super Bowl).
- A reasonable number of missing observations or large outliers.
- Historical trend changes, for instance due to product launches or logging changes.
- Trends that are non-linear growth curves, where a trend hits a natural limit or saturates [10].

Prophet model uses a decomposable time series model with three main components, trend, seasonality, and holidays, to predict $y(t)$. combined as:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

Where $g(t)$ is the trend function, $s(t)$ the seasonal changes term, and $h(t)$ is the term that represents holidays, the error term is represented by ε_t and is expected to be normally distributed [10].

4.1. Trend Model

The main component of a Prophet forecast is the trend $g(t)$, it defines how the time series population developed and is expected to continue. There are two possible choices for the trend function depending on data characteristics, a Nonlinear Saturating Growth model and a Linear Model [10].

4.2. Nonlinear Saturating Model

The growth of an indicator in this case is expected to saturate, and where this growth is not linear it saturates at carrying capacity [10].

$$g(t) = \frac{C}{1 + \exp(-k(t - m))}$$

Where C is the carrying capacity or saturation, k is the growth rate, t is the observation point and m an offset parameter.

Extra complexity is added when capacity C fluctuate in the time series, making C a function $C(t)$, such change impacts on the $g(t)$ equation:

$$g(t) = \frac{C(t)}{1 + \exp(-(k + a(t)^{T\delta})(t - (m + a(t)^{T\gamma}))}$$

Where rate at time t transforms in $k + a(t)^{T\delta}$, being δ a vector of rate adjustments, γ change point adjustment vector and T is a variance inferred from the data

4.3. Linear Trend with Change Point

When Forecasts do not saturate growth, and rate of growth is stable, the trend function $g(t)$ is:

$$g(t) = (k + a(t)^{T\delta}) t + (m + a(t)^{T\gamma})$$

Where k is the growth rate, δ is the rate adjustment, γ is the continuous change point adjustment [10]

4.4. Seasonality

It considers data patterns due to similar behaviour over data intervals. For example, the reason A&E department of a hospital is always busier between 5pm and 8pm is a result of people suffering injuries on commuting back home, taking relatives to the hospital after working hours, injuries when exercising, etc. There are longer and nested trends such as A&E department of a hospitals been busier on winter time than in summer due to respiratory illness, slippery conditions.

Prophet employs Fourier series to output a periodic effects model [10]. The seasonal effect function $s(t)$ is:

$$s(t) = \sum_{n=1}^N a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right)$$

Where a and b are observed seasonality, P is a regular period, N the number of observations.

4.5. Holidays and Events

Holidays and events provide a relatively large, and normally predictable changes in time series, frequently do not have a periodic pattern, thus the effects are not well factored by a smooth cycle. Such variations may affect positively or negatively the value of an KPI in timeseries depending on their characteristics. Prophet offer an extra variable in the model to feed a custom list of past and future holidays and events, and also assume they are independent [10]. Each entry i let D_L be set of past and future. Each holiday has a parameter k_i that corresponds for the change in the forecast.

$$Z(t) = [1(t \in D_1), \dots, 1(t \in D_L)]$$

Subsequently:

$$h(t) = Z(t)_{k_i}$$

5. DATA DESCRIPTION

This section aims to present data details, problems and reasons for which it has been used in this study in the comparison of the prediction models for ED.

5.1. Data Source

Choosing KPIs (Key Performance Indicators) in ED that explain pressure, is an inter-disciplinary matter. Input from healthcare professionals, managers and ultimately data scientists were employed in this study, the KPIs in analysis throughout this paper are Wait to be Seen Major in ED, Number of Attendances Major in ED, Unallocated Patients in ED with a DTA and Number of Beds Available on Medical Acute Unit. This data has been kindly provided by MFT (Medway Foundation Trust), and regards to numbers in MFT ED Acute Unit (hospital) located in Kent, South East England. For detailed description of each KPI and corresponding units (see Table 1).

“The NHS Constitution sets out that a minimum of 95% of patients attending an A&E department should be admitted, transferred or discharged within 4 hours of their arrival. In December 2016, NHS Improvement announced that hospitals will soon be rated against a new A&E standard. The 95% target will remain a key performance indicator, but it will be combined with clinical standards and staff and patient experience metrics”, [15].

The objective of this study also involves identifying pressure and optimising KPIs above mentioned, to provide results that may better care of patients and cost-effective solutions.

Table 1. Key Performance Indicators Description

Key Performance Indicator	Abbreviation	Description
Wait to be Seen Major in ED	WTBS	Average Number of minutes patients waited to be attended in the ED in each hour of the day
Number of Attendances Major in ED	NAED	Number of patients with major injuries attended in ED in each hour of the day
Unallocated Patients in ED with a DTA	UDTA	Number of patients attended in ED with a DTA (Decision to admit) waiting to be admitted into the hospital, each hour of the day
Number of Beds Available on Medical Acute Unit	NBAM	Number of beds available for patients attended in ED that need to be admitted into the hospital for continued care, each hour of the day

6. RESULTS

Simulations with both models have been performed to compare the accuracy in each of the described KPIs using both models (see Figure 2, Figure 3, Figure 4 and Figure 5). ARIMA is used as benchmark, once is the traditional and reliable model that is the standard in time series predictions in healthcare [3].

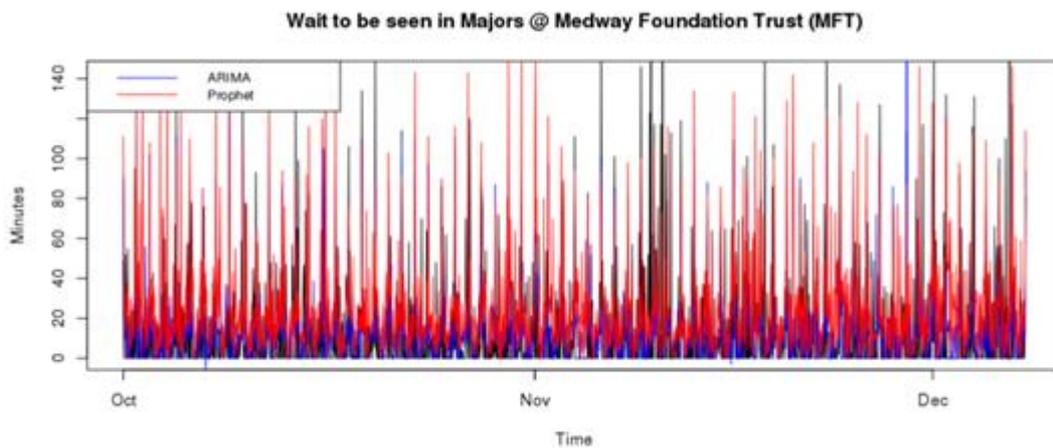


Figure 2. WTBS Actuals against ARIMA and Prophet predictions

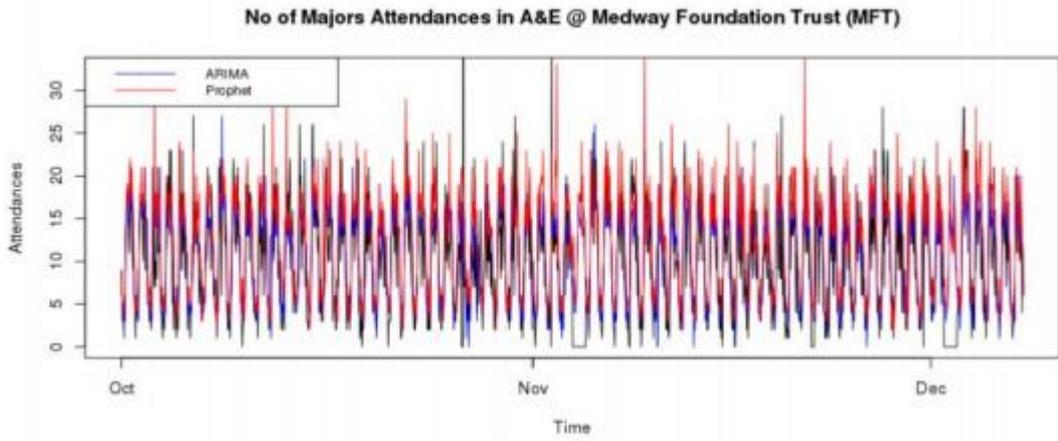


Figure 3. NAED Actuals against ARIMA and Prophet predictions

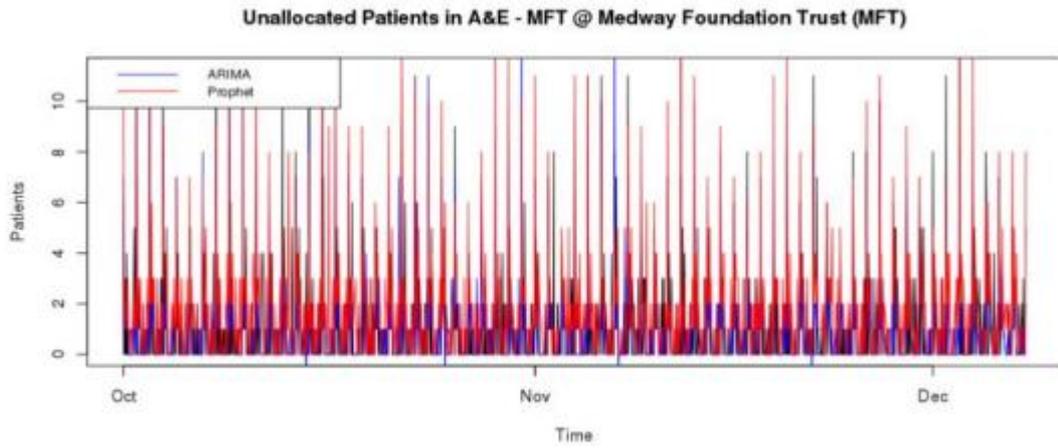


Figure 4. UDTA Actuals against ARIMA and Prophet predictions

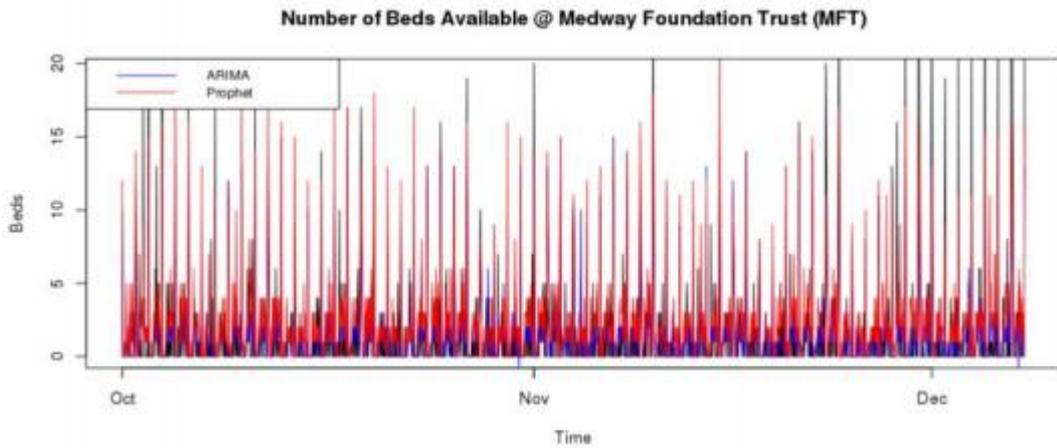


Figure 5. NBAM Actuals against ARIMA and Prophet predictions

6.1. Residual Analysis

The data that was made available to be analysed in this effort comprehends KPI values from 1 October 2018 to 2019-02-01. Different portions of the date have been analysed to remove possible bias, this training data has been separated in chunks of one month, the residual analysis metric of choice is RSME (see Table 2).

	01/10/2018 to 31/10/2018	01/11/2018 to 30/11/2018	01/12/2018 to 31/12/2018	01/01/2019 to 31/01/2019
WTBS	RMSE			
ARIMA	18.841882	18.851800	19.373623	19.648882
Prophet	23.690576	20.489990	18.342339	16.449243
Best Fit	ARIMA	ARIMA	Prophet	Prophet
NAED	RMSE			
ARIMA	5.0892403	5.0607747	5.0547193	5.0978627
Prophet	5.4374793	4.7028784	4.1900963	3.6906356
Best Fit	ARIMA	Prophet	Prophet	Prophet
UDTA	RMSE			
ARIMA	1.6292612	1.5961473	1.6238842	1.6880334
Prophet	1.8085393	1.5642065	1.3993058	1.2446480
Best Fit	ARIMA	Prophet	Prophet	Prophet
NBAM	RMSE			
ARIMA	2.2810902	2.2873253	2.4557430	2.6349061
Prophet	2.5716518	2.2242229	1.9978804	1.7820960
Best Fit	ARIMA	Prophet	Prophet	Prophet

7. CONCLUSIONS, LIMITATIONS AND FUTURE WORK

The prediction of healthcare emergency department key performance indicators in the emergency department of Medway Foundation Trust can be improved by using a new model such as Prophet. The residual analysis of the data point that out of 16 different simulations using both models, Prophet was the best fit in 11, whilst ARIMA could better explain future variables in 5 occasions (see Table 2).

Such analysis suggests that the use of a Prophet based model can improve decision making in approximately 69% of the cases if compared to the traditional ARIMA model in place. It is important to point that the difference of the RMSE did not exceeded 15% in any simulation, this

margin indicates that in the 5 occasions ARIMA is the best fit, Prophet prediction was within a reasonable reliability margin, however despite the accuracy gain, it also indicates such benefit is not extreme, and further studies and simulations with other datasets are necessary to further correlations.

Although the overall results suggest that generally the Prophet model explain forecasted variables with enhanced accuracy for the data being analysed, they also suggest limitations. When analysing data without a strong seasonal factor, likewise predicted data of October 2018, residuals tend to increase, in such case the ARIMA model often present better results (see Table 2). Another limitation was found when trying to analyse data with reduced quality, containing outliers likely to be wrong readings of actual values, it was observed that the Prophet model produced more outliers than the ARIMA model.

Future work involves exploring a hybrid method, mixing a number of averaging-based statistical forecast methods and trend based RNNs into a single time series prediction relying on classification of residuals of short-term and seasonal moments.

ACKNOWLEDGEMENTS

The authors would like to thank the Medway Foundation Trust for providing data, Transforming Systems and University of Greenwich to provide the platform to develop the research.

REFERENCES

- [1] L. Luo, L. Luo, X. Zhang, and X. He, "Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models."
- [2] A. E. Permanasari, I. Hidayah, and I. A. Bustoni, "SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence," 2013 Int. Conf. Inf. Technol. Electr. Eng., no. October, pp. 203–207, 2013.
- [3] Y. Sun, B. H. Heng, Y. T. Seow, and E. Seow, "Forecasting daily attendances at an emergency department to aid resource planning," BMC Emerg. Med., 2009.
- [4] S. S. Jones et al., "A multivariate time series approach to modeling and forecasting demand in the emergency department," J. Biomed. Inform., vol. 42, no. 1, pp. 123–139, 2009.
- [5] R. Kunst and K. Neusser, "A forecasting comparison of some var techniques," Int. J. Forecast., vol. 2, no. 4, pp. 447–456, Jan. 1986.
- [6] P. Aboagye-Sarfo, Q. Mai, F. M. Sanfilippo, D. B. Preen, L. M. Stewart, and D. M. Fatovich, "A comparison of multivariate and univariate time series approaches to modelling and forecasting emergency department demand in Western Australia," J. Biomed. Inform., vol. 57, pp. 62–73, 2015.
- [7] T. Zhu, L. Luo, X. Zhang, Y. Shi, and W. Shen, "Time-Series Approaches for Forecasting the Number of Hospital Daily Discharged Inpatients," IEEE J. Biomed. Heal. Informatics, vol. 21, no. 2, pp. 515–526, Mar. 2017.
- [8] M. Afilal, F. Yalaoui, F. Dugardin, L. Amodeo, D. Laplanche, and P. Blua, "Emergency department flow: A new practical patients classification and forecasting daily attendance," IFAC-PapersOnLine, vol. 49, no. 12, pp. 721–726, 2016.
- [9] P. C. Milner, "Ten-year follow-up of arima forecasts of attendances at accident and emergency departments in the Trent region," Stat. Med., vol. 16, no. 18, pp. 2117–2125, 1997.

- [10] S. J. Taylor and B. Letham, "Forecasting at Scale," PeerJ Prepr., vol. 5, no. e3190v2, p. 25, 2017.
- [11] Yi, Bu, and Kim, "An Enhanced Algorithm of RNN Using Trend in Time-Series," Symmetry (Basel), vol. 11, no. 7, p. 912, 2019.
- [12] A. Elliot and C. H. Hsu, "Time Series Prediction : Predicting Stock Price," no. 2, 2017.
- [13] N. Channouf, P. L'Ecuyer, A. Ingolfsson, and A. N. Avramidis, "The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta," Health Care Manag. Sci., 2007.
- [14] W. Vandaele, Applied Time Series and Box-Jenkins Models, 2nd ed. San Diego: Academic Press, Inc., 1983.
- [15] E. Fisher, N. Cooke O 'dowd, H. Dorning, E. Keeble, and L. Kossarova, "Quality at a cost," 2016.

AUTHORS

Diego Duarte works at Transforming Systems for 6 years, his currently position is Head of Digital Services, he also is a PhD Student at University of Greenwich, researching about predictive algorithms for time series based key performance indicators in healthcare. In his professional career before starting at Transforming Systems, Diego spent six years as IT Specialist in IBM Latin America and 3 years as Senior Software Engineer in Borland Latin America, amongst others. In his academic biography previous to the PhD, Diego completed a BSc in Computing Science at Universidade Catolica, an MBA at Fundacao Getulio Vargas and a MSc in Web Technology at University of Greenwich. Diego enjoys trying new ideas involving IoT, always with intent to apply technology to everyday life.



Julio Faerman is a software engineer working on technical enablement at Nubego, he collaborates with a great team to help builders worldwide, big or small, from classroom to production. Before that Julio is proud to have served for years in his favorite tech companies, including Amazon Web Services, Red Hat, Borland and many more. Julio is thankful for being a member of the many related technical communities, user groups, conferences and publications. His academic bio involves a BSc Computer Science at Universidade Federal de Uberlandia, an MBA at Fundacao Getulio Vargas, and several AWS Certifications. Now he is investing in saving the lives of cyclists with innovative IoT solutions being researched in an MSc IoT for eHealth, at Universidade Autonoma de Barcelona.

