FORECASTING FACEBOOK USER ENGAGEMENT USING HYBRID PROPHET AND LONG SHORT-TERM MEMORY MODEL

Kong Yih Hern[1], Lim Khai Yin[2] and Chin Wan Yoke[1*]

[1] Faculty of Computing and Information Technology, Tunku Abdul Rahman University College, Kampus Utama, Jalan Genting Kelang, 53300, Wilayah Persekutuan Kuala Lumpur, Malaysia
[2] Department of Computing and Information Technology, Tunku Abdul Rahman University College, Penang Branch Campus, Malaysia
*Corresponding author: chinwy@tarc.edu.my

ABSTRACT

Business forecasting remains a popular topic these days. A reliable business forecast often plays a vital part in an advertising campaign. The amount of attention acquired by posting an advertisement is one of the most essential criteria in determining the efficacy of the advertisement. The number of times that public users engage with a content signifies the amount of attention received, was measured by user engagement. With a good forecast, the advertisement could be promoted to a larger number of people. Facebook, as the most popular social media site, is favoured by majority of the advertisers. Therefore, this study addresses Facebook user engagement by forecasting the optimum date to post an advertisement. Different forecasting models, each with its own strengths and weaknesses, are used to model time series data with various properties. The objective of this study is twofold: to investigate the accuracy of the proposed Hybrid Prophet-LSTM that combines Long Short-Term Memory (LSTM) and FBProphet (Prophet) and to study the holiday impact on user engagement forecasting on Facebook brand pages. Data from 3 popular brand pages in the period of June 2018 to March 2019 was used in the experiments. The results show that the proposed hybrid model outperforms both the standalone LSTM and Prophet across the datasets. Besides, it is found that holiday effect could generally increase forecast accuracy. The optimum date for an advertisement campaign can therefore be determined based on the most forecasted user engagement, which consequently enhances the business income.

Keywords: Time Series forecasting, Hybrid forecasting, Business forecasting, Prophet, LSTM, Holiday effect

1.0 INTRODUCTION

User engagement refers to the attention, interactivity, perceived user control, and impression from the public users (Brien and Toms, 2008). Businesses are constantly seeking for innovative ways to improve the effectiveness of their advertisements. Adopting ineffective marketing techniques for a marketing campaign would not only squander corporate resources, but will also fail to get the desired results. The absence of user engagement with the advertisement platform was the most common reason for advertising campaigns underperforming (Goldsmith and Lafferty, 2002), (Frolova, 2014). The forecasted user engagement on a given advertisement plays an essential role in maximising the effect of an advertisement. According to the study by Frolova (2014), an effective advertisement can considerably raise volume of sales profits, foster consuming culture, fulfil customer wants for goods, and link advertiser and consumer audience in terms of communication channels. In other words, businesses should promote at the best time possible to achieve the most responses or user engagement from the audience.

Massive volumes of data from a large number of consumers are being collected through the media, particularly social media. A variety of studies by Schoen *et al.* (2013); Srinivasan *et al.* (2013); Breitenecker (2014); Kundi *et al.* (2014); Yasuko, Etuso and Akira (2014); Li *et al.* (2015); Di Gangi and Wasko (2016); Lee, Shia and Huh (2016); Debreceny (2019) have utilized social media data for various analysis. Researchers can study human behaviour patterns and predict user engagement using data from social media. In recent years, business forecasting has been a popular topic of study. The approach has been used to forecast time series data such as future stock movement (Sidi ,2020), traffic matrix (Azzouni and Pujolle, 2017), insurgency movement direction (Waeto, Chuarkham and Intarasit, 2017), and user engagement (Srinivasan *et al.* ,2013).

With an accurate forecasting result, user engagement for an advertisement may be easily attained. Selecting the right forecasting model is, therefore, of utmost importance. A variety of forecasting methods are utilised by businesses today. In this study, forecasting experiments are conducted by using Facebook data. This research employs the proposed Hybrid Prophet-LSTM by Kong, Lim and Chin (2021) to forecast user engagement that would in turn assist businesses in making managerial decisions on the commencement of an advertising campaign

In Section 2, applications of forecasting techniques in various fields are presented, showing how forecasting models are being used to solve various business problem. In Section 3, the details of dataset and proposed model are explained. The results of the evaluation can be found in Section 4. Section 5 discusses the findings and conclusion for this study

2.0 LITERATURE REVIEW

2.1 Forecasting in Businesses

Different models have been employed to analyse and solve various business problems (Polat, 2007). It is important to determine which model to use for solving a business problem.

Prediction and forecasting have recently been a popular topic. To overcome the network traffic problem, the study by Azzouni and Pujolle (2017) used forecasting model to predict network traffic matrix. Real-world data from the GEANT organisation network was used to test the feasibility of the forecasting model. The forecasting model is validated that could accurately predict traffic metrics. This forecast result is used to assist network operators in making decisions such as traffic accounting, short-time traffic scheduling, traffic rerouting, network design, long-term capacity planning, and network anomaly detection based on actual network traffic flows. The paper demonstrates how a forecasting model can be used to solve a business prediction problem by providing estimated future values that can be utilised to help making the decisions.

Yenidogan *et al.* (2018) used forecasting to tackle the Bitcoin forecasting difficulty in a recent study. The dataset contains two years' amount of Bitcoin exchange rates against a variety of currencies. The author employed a forecasting model to project future Bitcoin values, which is a critical subject for profit-seeking investors. The Bitcoin values were considered successful for future 90-days forecast with a precision of 94.5%. A credible forecast of future Bitcoin values would be valuable information for investors profiting from their Bitcoin investments.

Another study by Li *et al.* (2015) used forecasting to address a Twitter advertising problem. The click-through rate (CTR) on the Twitter timeline was forecasted using pointwise learning, pairwise learning, and further improvement version based on these two models. In the work, the authors proposed a model that used improvised algorithm based on pairwise and pointwise learning to learn user impressions with the click probability. The forecast outcome will alter how Twitter displays advertisements to users, leading to a greater CTR from Twitter users. The outcome from the model is found to be a more successful approach than traditional computational advertising, which are sponsored search and contextual advertising. The author concluded that the proposed method could significantly enhance the users' CTR on Twitter's advertisements.

The forecasting technique could also be used on social media data for a variety of purposes. Schoen *et al.* (2013) forecasted future events and developments using social media data. The events include area of politics, finance, entertainment, market demands, health, and other. The same study by Schoen *et al.* (2013) included influenza incidence, product sales, stock market movement, and electoral results as examples of forecasting applications using social media data.

As a result, user engagement is forecasted in order to decide the optimal date to promote. A reliable forecast of user engagement could aid businesses in making strategic decisions about how to execute a successful advertisement campaign that reaches the greatest number of people.

3.0 PROPOSED METHODOLOGY

3.1 Time series data

Facebook is the largest and the most favoured social media platform for public users (Bashar, Ahmad and Wasiq, 2012). To determine which variable is important to the research, the Facebook page and post metrics (*Insight - Pages*, 2021) were examined. Three brands in the categories of food, beverages, and cosmetics were chosen arbitrarily. The purpose of the following sections is to forecast the daily engagement received by a certain Facebook Page in order to reach the largest number of people possible. The target variable is the daily page engagement attribute, which was crawled from Facebook. Page engagement is a daily metric derived from user actions such as clicks, responses, comments, shares, and other forms of interaction with the page. Only one variable, customer page engagement, was examined and forecasted in this study.

Three datasets from two distinct sectors were gathered. Two years of daily time series data, starting on June 1, 2018, and ending on March 31, 2021, were gathered as a dataset from the three specified pages. Malaysia Public Holiday has been included to Prophet's holiday component. The purpose of this holiday dataset is to investigate the impact of holiday effects on time series forecast results. To assess the influence of holiday effects, a comparison study is carried out.

The number of times of users who engage with a certain page on a daily basis is referred to customer page engagement. As a result, this variable is a daily data variable with a daily count of user engagement. Users' clicks, reactions, shares, comments, and other actions are used to calculate user engagement. Overall, customer page engagement is a measurement of how much public users pay attention to a page.

3.2 Proposed model

This study uses the proposed Hybrid Prophet-LSTM by Kong, Lim and Chin (2021) to enhance forecast accuracy. In the suggested hybrid methodology, Prophet is used as the linear model, while LSTM is used to address the residual nonlinear connection in the time series data. Prophet is used to model regular, irregular, and non-regular holiday events. Because time series data contains both linear and nonlinear structure, a nonlinear model, such as LSTM, is used to represent the residual matrix from a linear model. With its remarkable potential of addressing nonlinearity relationships in time series, LSTM is utilised to model nonlinear relationships in the residual matrix to produce better forecast results. To create the forecast result, the time series data were first fitted into Prophet. The residual matrix was computed using Prophet's forecasted output, and the residual matrix was then fitted into LSTM. The forecasted residual is used to compute hybrid forecast output. Finally, the output produced by the hybrid model is evaluated using several performance metrics and compared to the results of various models.

4.0 EMPIRICAL RESULTS

4.1 Time series decomposition

An analysis is performed to study the individual components in the time series data. The user engagement data was analysed using decomposition feature in Prophet. A decomposition result was generated to understand various characteristic in a time series such as its trend, holiday effects, and multiple forms of seasonality.

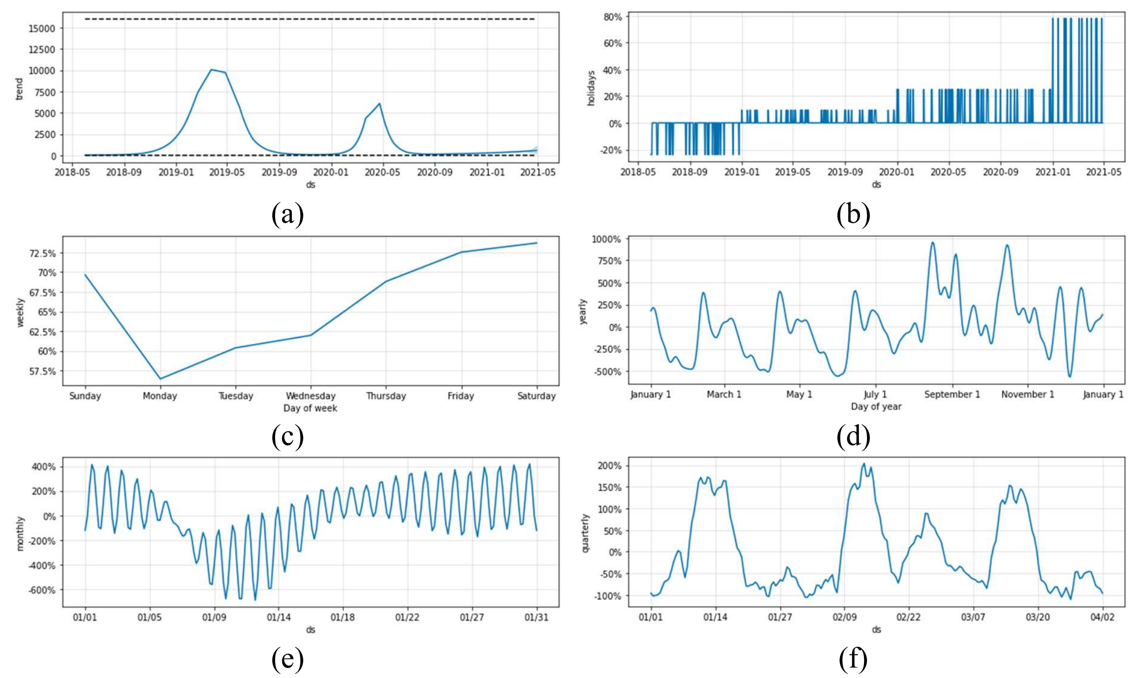

Figure 2. Example of individual components from Prophet’s decomposition result, a) Data trend, b) Holiday effects, c) Weekly seasonality, d) Yearly seasonality, e) Monthly seasonality, f) Quarterly seasonality

The decomposition result displayed in Figure 2 is created using Dataset 1. As shown in Figure 2(a), the engagement does have 2 growths in the historical data but then declined after few months. There is little evidence of continuous potential growth, and the trend remains in the range of 0 to 11,000 engagements.

In Figure 2(b), the holiday effects graph demonstrates that only the holiday events occurring before year 2019 have a negative impact on engagements. Other holiday-related events have a beneficial impact on the time series data. When there is a holiday event, engagement decreases before the year 2019, but increases in various scales in the year 2019 and later. Each of these holiday occurrences had an influence that was discovered and used to change the predicted estimate. Prophet managed the irregular holiday effects during modelling by applying this analysis result.

The weekly seasonality shown in Figure 2(c) demonstrates that user engagement to this page drops on Sunday, then gradually increases until Saturday, which is also the week's peak engagement. This explained the user engagement behaviour over the course of a week, demonstrating that they are more willing to engage with the page as the weekend approaches. The yearly seasonality is shown in Figure 2(d), and we can see a uniform yearly seasonality in this decomposition result. However, from September 2020 to November 2020, the model captures a lot of noise. These factors will have an impact on the modelling outcome, which will be inaccurate for the specific period in the yearly seasonality. Monthly seasonality is depicted in Figure 2(e) as a wave pattern, with engagement becoming stable at 0% at the beginning of the month, then progressively increasing in total, regaining the lost value until the conclusion of the month, but extremely instable at the same time. Weekly seasonality has an impact on monthly seasonality. As a result, we can refer to Figure 2(c) for the weekly seasonality decomposition explanation. The quarterly seasonality, which is the form of seasonality for one quarter, is shown in Figure 2(f). The quarterly seasonality rises from -100% at the start of each month to 150% in the 15th days of each month, then returning the value obtained earlier. Monthly and weekly seasonality have a significant impact on quarterly seasonality; see Figure 2(c) and Figure 2(e) for further information.

4.2 Standalone and Hybrid Prophet-LSTM algorithm

The linear relationship in time series data was fitted using Prophet and the remaining pattern under the Prophet residual was fitted into LSTM. Five distinct methods are compared for creating a reliable forecast result and examining the impact of the holiday effect on the forecast outcome. Prophet, Prophet without holiday, LSTM, Hybrid Prophet-LSTM, and Hybrid Prophet-LSTM without holiday were among the approaches used. These methods are evaluated using different performance metrics including Weighted Mean Absolute Percentage Error (WMAPE), R^2 score, Root mean square error (RMSE), and Mean Absolute Deviation (MAD).

Table 1. Performance metrics for Prophet, LSTM, and Hybrid models

		Prophet	Prophet (No Holiday)	LSTM	Hybrid Prophet-LSTM	Hybrid Prophet-LSTM (No Holiday)
Dataset 1	WMAPE	47.8679%	46.9547%	17.7578%	16.8768%	15.6264%
	R^2	99.9339%	99.9401%	95.8725%	99.9946%	99.9953%
	RMSE	763.306	744.791	347.279	243.283	227.217
	MAD	504.749	495.586	182.326	173.859	161.094
Dataset 2	WMAPE	21.1959%	21.4081%	6.1050%	5.8045%	5.5680%
	R^2	99.9930%	99.9928%	94.7999%	99.9995%	99.9996%
	RMSE	5088.143	5163.085	2283.617	1381.347	1298.541
	MAD	4037.848	4079.288	1160.590	935.641	855.583
Dataset 3	WMAPE	55.3595%	58.1690%	22.8360%	18.8923%	31.8675%
	R^2	99.5889%	99.4153%	58.5297%	99.9799%	99.8619%
	RMSE	328.844	337.710	783.955	112.955	283.801
	MAD	168.269	170.354	63.780	58.906	117.100

Table 1 compares the results of the three models with different approaches. Models 1 and 2 are the standalone Prophet and LSTM, and Model 3 is the Hybrid Prophet-LSTM. The models were also compared with and without the holiday component. Prophet has a WMAPE of 47.86 %, 21.19 %, and 55.35 %, respectively using the three datasets. By comparing the error rates of the standalone Prophet and LSTM models, the results show that Prophet's error rate is at least double or more than LSTM's error rate. LSTM outperforms Prophet by producing fewer errors and a lower overall error rate. When LSTM is compared to the Hybrid model, the hybrid model outperforms LSTM model in every aspect. By having reduced mistakes and error rates, as well as a higher R^2 value, the hybrid model outperforms the LSTM. The hybrid model is an alternative to the traditional model.

Prophet model did not demonstrate good modelling in this case since Prophet's holiday component does not significantly improve Prophet's performance. Despite the fact that holiday effects were not adequately visible in Prophet's forecasting results, the holiday component has a significant impact on the hybrid model forecast result for Dataset 3. The holiday component has a minor influence on Datasets 1 and 2, but has a considerable impact on Dataset 3 with an error reduction of 18.89 % to 31.86 % without the holiday component. In overall, the holiday component could improve the Hybrid Prophet-LSTM model in producing a more accurate forecast.

Table 2 shows that LSTM outperforms Prophet when it comes to modelling user engagement time series data. As a result, a hybrid model may model both linear and nonlinear time series data with a steady performance because it incorporates the strengths of both linear and nonlinear models.

During the study, it was discovered that although the Prophet linear model can detect seasonality in Dataset 1 and 3, but the seasonality captured is unusual and does not show an observable pattern, which can considerably increase forecast error. When there is no observable seasonality pattern in the time series, the forecasting accuracy for these datasets is relatively low.

The forecasting models demonstrate their feasibility modelling time series data to forecast user page engagement as a result of the findings. The hybrid model, which was shown to be the best, had a forecast error range of 5.80 % to 18.89 %. Businesses are able to forecast a page engagement by leveraging a good model, which can then be used to determine the optimum day to begin an advertising campaign. Posting an advertisement on a day with higher engagement indicates that the advertisement will reach a larger group of audience.

5.0 CONCLUSIONS

Hybrid Prophet-LSTM was utilised in this work to combine linear and nonlinear models to produce improved forecast results. The proposed approach incorporates features such as a customizable calendar of events or holidays. In addition to irregular holiday occurrences, this study models seasonality and trend components. Experiments were carried out to verify the effectiveness of the proposed model. We can see that without the holiday components in the hybrid model, the models perform better in overall. The suggested model has the smallest forecasting errors and performs well across a wide range of datasets and scales of variance. As a result, it can be inferred that while attempting to produce an accurate forecast, there are two critical factors to consider. The linear model's compatibility would be the first concern. The performance of the linear model was found to have a significant impact on the hybrid forecast result. In the hybrid model, a well-performed output in the linear model would offer an exceptional outcome. The selection of features is the second factor. Only one variable was chosen for forecasting in this study, resulting in a univariate analysis. These studies by Hummel and Sligo (1971); Saccenti *et al.* (2014) implied that both multivariate and univariate approaches should be used because the results from these two analyses are complementary. Analysing the relationship between the dependent variable and other independent variables would be the future work for this study. It can be stated that the results of this study will be possible to forecast dates with the most user engagement. However, any managerial judgment should not be made solely on the basis of this variable. This study serves the purpose of exploring the expected advertisement effect. Businesses should instead validate the result with numerous approaches, such as referring to the expert knowledge and experiences in the domain, conducting experiment using other datasets, and compare the result with other single-variable data analysis on advertising, as advertising investment is a complicated practice in the actual world (Dawes *et al.*, 2018).

REFERENCES

Azzouni, A. and Pujolle, G. (2017) ‘A Long Short-Term Memory Recurrent Neural Network Framework for Network Traffic Matrix Prediction’. Available at: http://arxiv.org/abs/1705.05690.

Bashar, A., Ahmad, I. and Wasiq, M. (2012) ‘Effectiveness of Social Media As a Marketing Tool: an Empirical Study’, *International Journal of Marketing, Financial Services & Management Research*, 1(11), pp. 88–99. Available at: www.indianresearchjournals.com.

Breitenecker, R. J. (2014) ‘Customer engagement behaviour in online social networks – the Facebook perspective Sofie Bitter * and Sonja Grabner-Kräuter’, 14.

Brien, H. L. O. and Toms, E. G. (2008) ‘What is User Engagement? A Conceptual Framework

for Defining User Engagement with Technology’, *Journal of the American Society for Information Science and Technology 59(6):938-955*, (April). doi: 10.1002/asi.20801.

Dawes, J. *et al.* (2018) ‘Forecasting advertising and media effects on sales: Econometrics and alternatives’, *International Journal of Market Research*, 60(6), pp. 611–620. doi: 10.1177/1470785318782871.

Debreceny, R. (2019) ‘Research in Social Media : Data Sources and Methodologies Research in Social Media : Data Sources and Methodologies’, (December 2017). doi: 10.2308/isys-51984.

Frolova, S. (2014) ‘Svetlana Frolova THE ROLE OF ADVERTISING IN PROMOTING A PRODUCT Thesis CENTRIA UNIVERSITY OF APPLIED SCIENCES Degree Programme in Industrial Management’, (May).

Di Gangi, P. M. and Wasko, M. (2016) ‘Social media engagement theory: Exploring the infuence of user engagement on social media usage’, *Journal of Organizational and End User Computing*, 28(2), pp. 53–73. doi: 10.4018/JOEUC.2016040104.

Goldsmith, R. E. and Lafferty, B. A. (2002) ‘Consumer response to Web sites and their influence on advertising effectiveness’, *Internet Research*, 12(4), pp. 318–328. doi: 10.1108/10662240210438407.

Hummel, T. J. and Sligo, J. R. (1971) ‘Empirical comparison of univariate and multivariate analysis of variance procedures’, *Psychological Bulletin*, 76(1), pp. 49–57. doi: 10.1037/h0031323.

Insight - Pages (2021). Available at:

https://developers.facebook.com/docs/platforminsights/page (Accessed: 29 March 2020).

Kong, Y. H., Lim, K. Y. and Chin, W. Y. (2021) ‘Time Series Forecasting using a hybrid Prophet and LSTM model’, in.

Kundi, F. M. *et al.* (2014) ‘Detection and scoring of Internet Slangs for sentiment analysis using SentiWordNet’, *Life Science Journal*, 11(9), pp. 66–72. doi: 10.6084/M9.FIGSHARE.1609621.

Lee, T.-S., Shia, B.-C. and Huh, C.-L. (2016) ‘Social Media Sentimental Analysis in Exhibition’s Visitor Engagement Prediction’, *American Journal of Industrial and Business Management*, 06(03), pp. 392–400. doi: 10.4236/ajibm.2016.63035.

Li, C. *et al.* (2015) ‘Click-through prediction for advertising in twitter timeline’, *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data*

Mining, 2015-Augus, pp. 1959–1968. doi: 10.1145/2783258.2788582.

Polat, C. (Niğde Ü. (2007) 'The Role of Forecasting and Its Potential for Functional Management: A Review from the Value-Chain Perspective', *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(1), pp. 373–398.

Saccenti, E. *et al.* (2014) 'Reflections on univariate and multivariate analysis of metabolomics
data', *Metabolomics*, 10(3), pp. 361–374. doi: 10.1007/s11306-013-0598-6.

Schoen, H. *et al.* (2013) 'The power of prediction with social media', *Internet Research*, 23(5),
pp. 528–543. doi: 10.1108/IntR-06-2013-0115.

Sidi, L. (2020) 'Improving S&P stock prediction with time series stock similarity'. Available at: http://arxiv.org/abs/2002.05784.

Srinivasan, B. V. *et al.* (2013) 'Will your facebook post be engaging?', *International Conference on Information and Knowledge Management, Proceedings*, pp. 25–28. doi: 10.1145/2512875.2512881.

Waeto, S., Chuarkham, K. and Intarasit, A. (2017) 'Forecasting time series movement direction
with hybrid methodology', *Journal of Probability and Statistics*, 2017. doi: 10.1155/2017/3174305.

Yasuko, K., Etuso, G. and Akira, I. (2014) 'P REDICT F ACEBOOK I MPRESSIONS ADOPTING A MATHEMATICAL MODEL OF THE H IT', 2(1), pp. 63–68.

Yenidogan, I. *et al.* (2018) 'Bitcoin Forecasting Using ARIMA and PROPHET', *UBMK 2018 - 3rd International Conference on Computer Science and Engineering*, (February 2019), pp. 621–624. doi: 10.1109/UBMK.2018.8566476.