FORECASTING FACEBOOK USER ENGAGEMENT USING HYBRID PROPHET LSTM AND IFOREST

(Date received: 24.01.2022/Date accepted: 08.04.2022)

Kong Yih Hern[1], Lim Khai Yin[2], Chin Wan Yoke[3]

[1,3] Tunku Abdul Rahman University College, Jalan Genting Kelang, Setapak, 53300 Kuala Lumpur, P.O. Box 10979, 50932 Kuala Lumpur, Malaysia

[2] Department of Computing and Information Technology, Tunku Abdul Rahman University College, Penang Branch Campus

E-mail: [1] kongyh-wa15@student.tarc.edu.my, [2] limky@tarc.edu.my, [3] chinwy@tarc.edu.my

ABSTRACT

Business forecasting remains a popular topic these days. A reliable business forecast often plays a vital role in an advertising campaign. The amount of attention acquired by posting an advertisement is one of the most essential criteria to determine the effectiveness of the advertisement. The number of times that public users engage with a content signifies the amount of attention received, which 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 preferred by the majority of 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 threefold: to investigate the accuracy of the proposed Hybrid Prophet-LSTM that combines Long Short-Term Memory (LSTM) and FBProphet (Prophet), to study the holiday impact on user engagement forecasting on Facebook brand pages, and to study the effect of implementing Isolation Forest (iForest) on the dataset and its contribution to the forecast result. Data from three popular brand pages were used in the experiments in the period of June 2018 to March 2021. 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. In general, datasets pre-processed using iForest can reduce the forecast error under specific conditions. Therefore, the optimum date for an advertisement campaign can be determined on the basis of the most anticipated 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 attention, interaction, perceived user control, and impression from public users (Brien and Toms, 2008). Businesses are constantly seeking 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 advertising platform was the most common reason 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), effective advertising can considerably increase volume of sales profits, foster consumption 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, Akira (2014); Li *et al*. 2015; Di Gangi and Wasko (2016); Lee *et al*., (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 (Waeto *et al*., 2017), and user engagement (Srinivasan *et al*., 2013). However, there are some uncertainties in exploring the datasets collected in each of these applications. In this paper, these uncertainties were removed using iForest before fitting the datasets into forecasting models. The objectives of resolving the uncertainties of the datasets are to decrease modelling error, hence increasing the forecast accuracy.

With an accurate forecasting result, user engagement for an advertisement may be easily attained. Selecting the right forecasting model is, therefore, of utmost importance. Today, businesses are using a variety of forecasting methods. In this study, forecasting experiments are conducted using Facebook data. This research employs the proposed Hybrid Prophet-LSTM by Kong *et al.,* (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 the data set and the 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 to solve 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 the LSTM model to predict the network traffic matrix. Real-world data from the GEANT organisation network was used to test the feasibility of the LSTM model. The LSTM 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.

Yenidogan *et al*. (2018) used forecasting to tackle the difficulty of Bitcoin forecasting in a recent study. The dataset contains two years of Bitcoin exchange rates against a variety of currencies. The author employed Prophet to project future Bitcoin values, which is a critical subject for profit-seeking investors. Bitcoin values were considered successful for the future 90-day forecast with a precision of 94.5%. A credible forecast of future Bitcoin values would be valuable information for investors who want to profit from their Bitcoin investments. The articles demonstrate how a forecasting model can be used to solve a business prediction problem by providing estimated future values that can be used to help make decisions.

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 a further improvement version based on these two models. In the work, the authors proposed a model that used an improvised algorithm based on pairwise and pointwise learning to learn user impressions with the click probability. The forecast result will alter how Twitter displays advertisements to users, leading to a higher 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) forecast future events and developments using social media data. The events include the area of politics, finance, entertainment, market demands, health, and others. 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 forecast to decide the optimal date to promote. A reliable forecast of user engagement could help companies make strategic decisions about how to execute a successful advertising campaign that reaches the largest number of people.

However, many of the researchers tend to neglect the effect of holiday events on the forecast result, as well as the complexity of forecasting human behaviour-related events such as user engagements in this study. To solve the above-mentioned business problem would require a hybrid methodology such as Hybrid LSTM-Prophet model (Kong *et al.,* 2021).

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. Businesses should validate the result from both multivariate and univariate approaches with numerous methods, such as referring to expert knowledge and experience in the domain, conducting experiments using other datasets, and comparing the result with other single-variable data analysis in advertising, as advertising investment is a complicated practice in the real world (Dawes *et al.,* 2018). However, this study would only be scoped in univariate analysis, or single-variable business forecasting.

3.0 PROPOSED METHODOLOGY

3.1 Time Series Data

The study by Bashar *et al.,* (2012) shows Facebook is the largest and the most favoured social media platform for public users. To determine which variable is important to the research, the Facebook page and post metrics (*Insight - Pages,* 2022) were examined. Three brands were arbitrarily chosen from each category of food, beverages and cosmetics. 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. In this study, only one variable, customer page engagement, was examined and forecasted.

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. Malaysian Public Holiday has been included in 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 users engage with a certain page on a daily basis is called 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. In general, customer page engagement is a measurement of how much public users pay attention to a page.

The datasets were examined to see whether any outliers existed. The term "outlier" refers to observations that differ from the majority of the data (Rousseeuw and Hubert, 2018). These abnormal observations are also known as anomalies in a machine learning context. However, Pollet and van der Meij (2017) said that outliers may be caused by various errors such as data entry errors, experimental errors, data processing errors, measurement errors, and natural. If the outliers were not produced by any of the above-mentioned errors, these observations should be categorised as the novelties of the data. Although we ensured that the datasets do not have any outliers caused by the mentioned errors, however we conducted experiments using outlier detection and removal technique to measure and evaluate the outcome of removing these abnormal observations from the dataset and its contribution to the forecast result.

3.2 Proposed Procedure

This study uses the proposed Hybrid Prophet-LSTM by Kong *et al.,* (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, and nonregular 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 to address 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 forecast residual is used to compute the 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.

The datasets were analysed using (iForest), the outlier detection technique that we selected in this study. iForest is proved to have great performance with high efficiency with only a very small numbers of tress or sub-sampling size (Liu *et al.,* 2012). Many researchers Liu *et al.,* (2012); Ding and Fei (2013); Hofmockel and Sax (2018); Gao *et al.* (2019); Holmer (2019); Hariri *et al.,* (2021) have been adapting iForest as a tool for anomaly detection and even creating an iForest-based approach. Liu *et al.,* (2012) said that iForest takes the advantages of anomaly properties where: the anomalies are the minority and consist of fewer instances; the anomalies have values that are very different from the normal instances. In a simpler word, anomalies are less and different when compared to the normal observations. The study by Liu *et al.,* (2012) demonstrates the algorithm of iForest to estimate s, the anomaly score of the data points. The anomaly score s is used to make the following assessment: normal common samples (s lower than 0.5), normal uncommon samples (s near 0.5), and outliers (s very close to 1).

In this study, it is observed that many unexplainable data points occurred in the time series. These data points were further analysed and it is observed that these data points were not explainable by the time series trend, seasonality, nor holiday effect. Additionally, these data points were not caused by either of the errors. Therefore, in this study, these anomalies were identified and removed using iForest, to evaluate the effect of iForest on the datasets and its impact on the forecast result.

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 characteristics of a time series such as its trend, holiday effects, and multiple forms of seasonality.

In order to achieve a reliable forecast, it is important to capture the components in the time series. Simpler to say, the clearer the captured pattern, the better the model understands the data. The decomposition result shown in Fig. 1 is created using Dataset 1. As shown in Fig. 1, the Prophet model decomposes the time series into individual components and analyses each of these components separately. Parameters for each of these components can also be adjusted individually. Each of these graphs tells information for each component in the time series. Fig. 1(a) shows the growth trend of engagement, Fig. 1(b) shows the impact scale of holiday effects on engagement, Fig. 1(c) shows the weekly seasonality, Fig. 1(d) shows the yearly seasonality, Fig. 1(e) shows the monthly seasonality and Fig. 1(f) demonstrates the quarterly seasonality.

In Fig. 2, it shows the decomposition analysis result after the dataset is pre-processed using iForest. As the 'outliers' were detected and removed during iForest, the dataset now does not have the 'peak points' that differ from most common observations. Here we demonstrate the difference before and after using iForest. In Fig. 2(a), the growth trend reduced in total variance. The pattern of different forms of seasonality in Fig. 2(c), Fig. 2(d), Fig. 2(e), and Fig. 2(f) changed accordingly. Furthermore, in Fig. 2(d), the yearly seasonality pattern become more stably recursive compared to Fig. 1(d) that captured some noise in the time series. Although Prophet's model seems to be able to capture the pattern of the series better than using the original data, here we only compare the effect of iForest with the time series, and the contribution of iForest to the forecast result will be discussed later.

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).

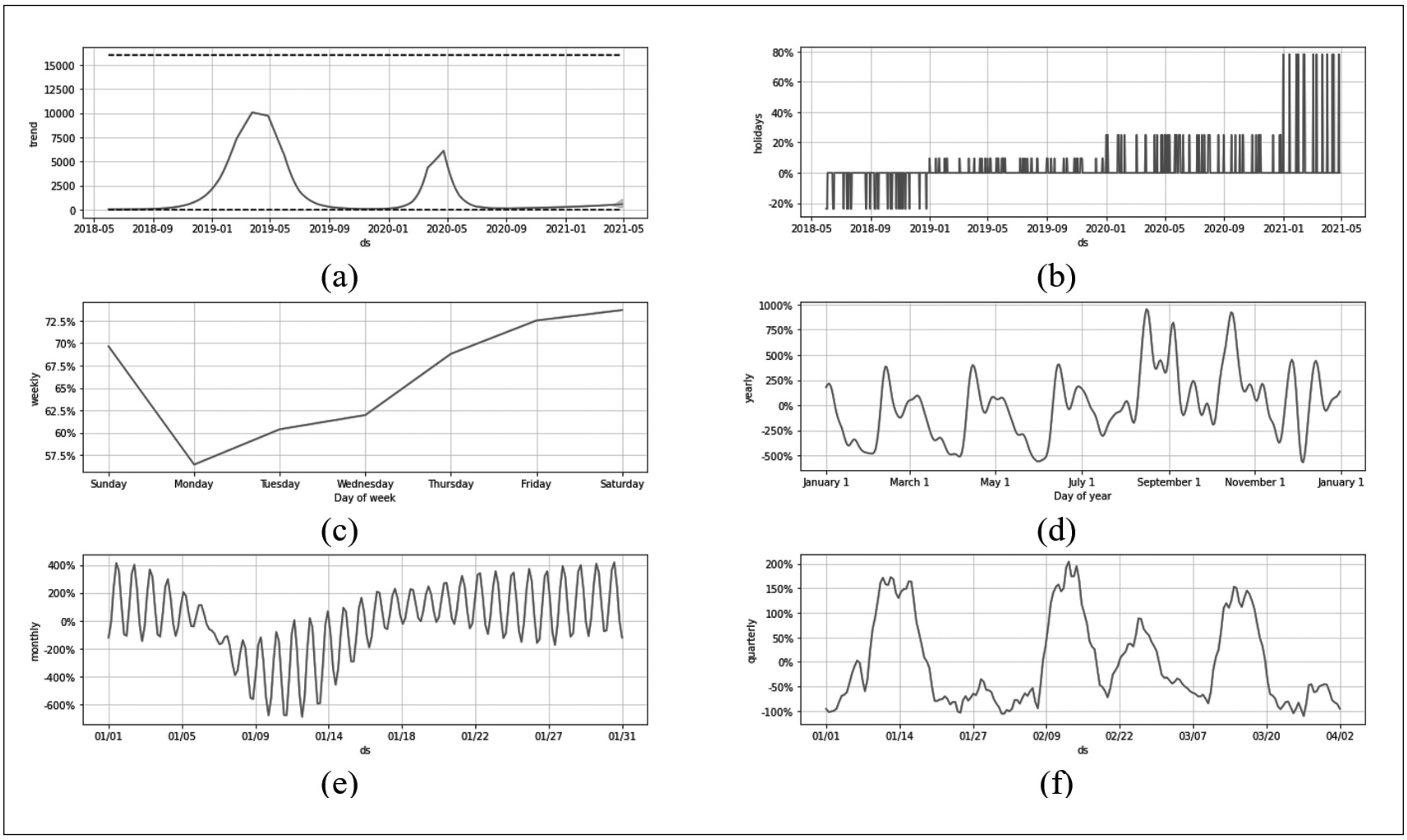

Figure 1: Individual components from Prophet's decomposition result
a) Data trend, b) Holiday effects, c) Weekly seasonality, d) Yearly seasonality, e) Monthly seasonality, f) Quarterly seasonality

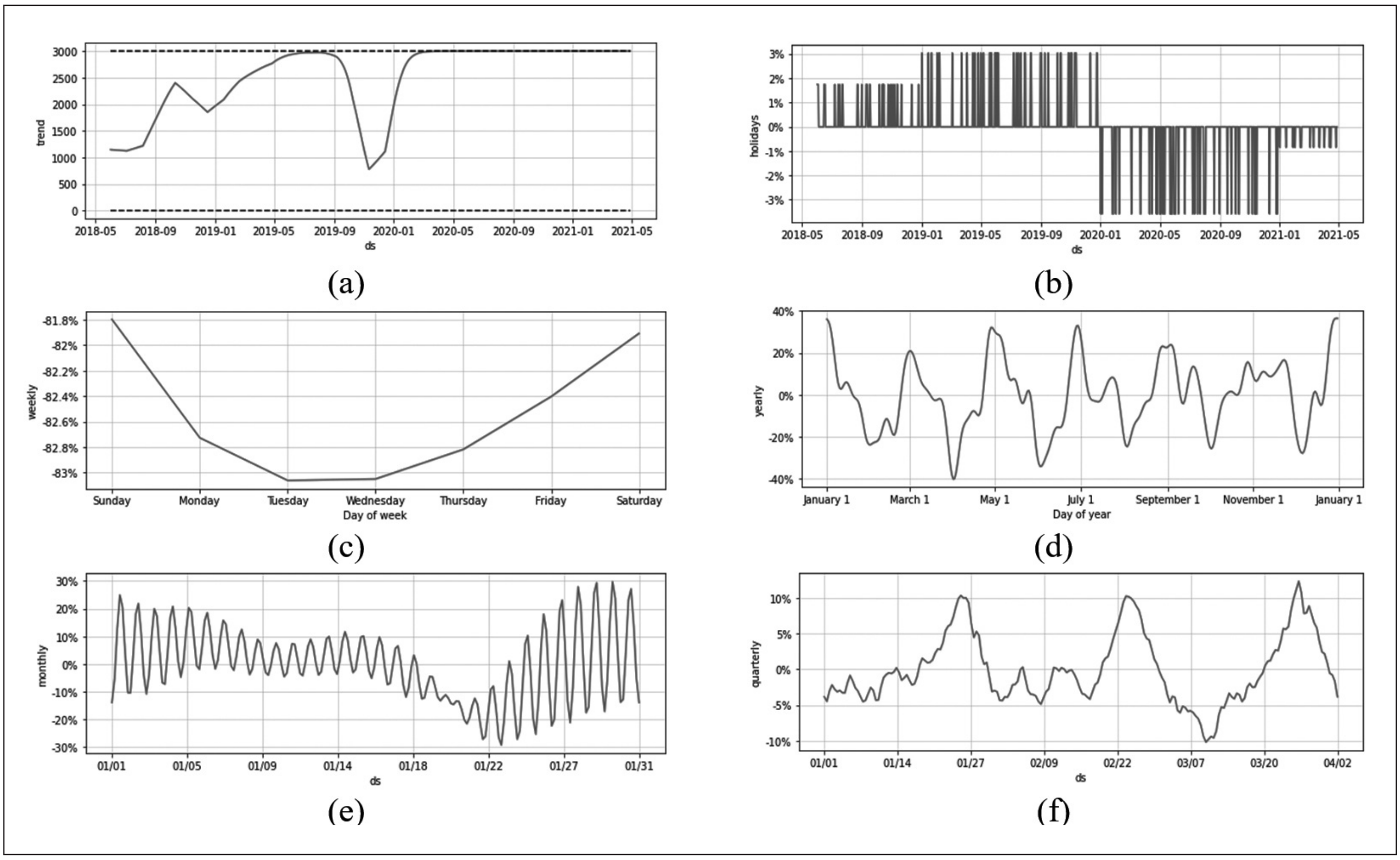

Figure 2: Individual components of Prophet's decomposition result after iForest
a) Data trend, b) Holiday effects, c) Weekly seasonality, d) Yearly seasonality, e) Monthly seasonality, f) Quarterly seasonality

Table 1 compares the results of the three models with different approaches without removing anomalies using iForest. Models 1 and 2 are the standalone Prophet and LSTM, and Model 3 is the Hybrid Prophet-LSTM. 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 the 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. Overall, the holiday component could improve the hybrid Prophet-LSTM model in producing a more accurate forecast.

Table 1 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 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 Datasets 1 and 3, 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 can forecast page engagement using a good model, which can then be used to determine the optimal day to start an advertising campaign. Posting an advertisement on a day with higher engagement indicates that the advertisement will reach a larger group of audience.

Similarly, Table 2 uses datasets that were pre-processed using iForest. Prophet has a WMAPE of 45.28 %, 5.60 %, and 18.11 %, respectively, using the three datasets as shown in Table 1. Compared to the result shown in Table 1, Prophet generally performs better for the three datasets. It is observed that iForest can generally increase Prophet and LSTM's performance whether with holiday component included or not.

However, it is found that the Hybrid model performs terribly bad with the highest forecast error and highest RMSE and MAD in Dataset 1. The forecast error for Prophet remains the same and lower forecast error for LSTM. It is found that huge forecast errors occurred while modelling the residual matrix in the Hybrid Prophet-LSTM modelling process. Although when modelling the pre-processed Dataset 1 has a WMAPE of 45.27%, Hybrid Prophet-LSTM showed that it has the capability to reduce error even if there was high forecast error occurred when modelling using Prophet in creating residual matrix. Therefore, this can be explained by the iForest outlier removal procedure. As mentioned earlier, there were no outliers caused by any kind of errors. The reason why iForest is conducted is to experiment the effect of removing the "outlier" of the dataset and its contribution to the forecast accuracy. In Table 2, iForest generally increases the accuracy of the forecast by reducing error except for the data set using the hybrid Prophet-LSTM model in Dataset 1. This result is tally with the modelling result because both Prophet and LSTM demonstrated to capture a clearer pattern of the series than before. For the huge error that occurred in Dataset 1 Hybrid Prophet-LSTM, it can be explained that LSTM failed to model the residual matrix after removing the natural of the dataset, provided there are no outliers in the data.

Table 1: Performance metrics for Prophet, LSTM, and Hybrid models without using iForest

		Prophet	Prophet (No Holiday)	LSTM	Hybrid Prophet-LSTM	Hybrid Prophet-LSTM (no holidays)
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 2: Performance metrics for Prophet, LSTM, and Hybrid models using iForest

		Prophet	Prophet (No Holiday)	LSTM	Hybrid Prophet-LSTM	Hybrid Prophet-LSTM (no holidays)
Dataset 1	WMAPE	45.2756%	48.3359%	3.6556%	86.3346%	93.6136%
	R^2	43.8439%	38.2864%	99.1561%	0.1821%	3.1823%
	RMSE	242.941	254.679	29.793	464.702	493.059
	MAD	195.043	208.227	15.762	373.262	404.732
Dataset 2	WMAPE	5.6033%	5.6449%	1.1806%	1.1026%	1.0905%
	R^2	73.1712%	72.7941%	85.0587%	99.5299%	99.5470%
	RMSE	1192.951	1201.306	911.127	228.613	224.984
	MAD	947.188	954.237	199.254	130.310	115.753
Dataset 3	WMAPE	18.1109%	18.2126%	4.1066%	3.7666%	3.6554%
	R^2	71.5381%	71.2029%	96.6484%	99.1497%	99.2107%
	RMSE	8.431	8.480	2.899	1.770	1.716
	MAD	6.524	6.560	1.476	1.636	1.778

The proposed approach incorporates features such as a customizable calendar of events or holidays. In addition to irregular holiday events, this study models seasonality and trend components. Experiments were carried out to verify the effectiveness of the proposed model. We can see that by including the holiday components in the hybrid model, the models perform better overall. The suggested model has the smallest forecast errors and performs well on a wide range of datasets and scales of variance. Additionally, iForest could significantly reduce the forecast error when the characteristic of the dataset is well defined. It is best to seek for advice from business expert in the respective domains.

5.0 CONCLUSIONS

The Hybrid Prophet-LSTM was utilised in this work to combine linear and nonlinear models to produce improved forecast results. As a result, it can be inferred that while attempting to produce an accurate forecast, there are two critical factors to consider. The compatibility of the linear model 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-performing output in the linear model would offer an exceptional outcome. The selection of features is the second factor. In this study, only one variable was chosen for forecasting, which resulted in a univariate analysis. Analysing the relationship between the dependent variable and other independent variables would be 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 effect of advertising. Businesses should instead validate the result with numerous approaches, such as referring to expert knowledge and experience in the domain, conducting experiments using other datasets, and comparing the result with other single-variable data analysis in advertising, as advertising investment is a complicated practice in the real world. ■

REFERENCES

[1] 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.

[2] 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.

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

[4] 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.

[5] 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.

[6] 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.

[7] Ding, Z. and Fei, M. (2013) An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window, IFAC Proceedings Volumes (IFAC-PapersOnline). IFAC. doi: 10.3182/20130902-3-CN-3020.00044.

[8] 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).

[9] 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.

[10] Gao, R. et al. (2019) 'Research and Improvement of Isolation Forest in Detection of Local Anomaly Points', Journal of Physics: Conference Series, 1237(5). doi: 10.1088/1742-6596/1237/5/052023.

[11] 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.

[12] Hariri, S., Kind, M. C. and Brunner, R. J. (2021) 'Extended Isolation Forest', IEEE Transactions on Knowledge and Data Engineering, 33(4), pp. 1479–1489. doi: 10.1109/TKDE.2019.2947676.

[13] Hofmockel, J. and Sax, E. (2018) 'Isolation forest for anomaly detection in raw vehicle sensor data', VEHITS 2018 - Proceedings of the 4th International Conference on Vehicle Technology and Intelligent Transport Systems, 2018-March(Vehits 2018), pp. 411–417. doi: 10.5220/0006758004110416.

[14] Holmer, V. (2019) 'Hybrid Extended Isolation Forest'.

[15] 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.

[16] Insight - Pages (2022). Available at: https://developers.facebook.com/docs/platforminsights/page (Accessed: 29 March 2020).

[17] Kong, Y. H., Lim, K. Y. and Chin, W. Y. (2021) 'Time Series Forecasting Using a Hybrid Prophet and Long Short-Term Memory Model', Communications in Computer and Information Science, 1489 CCIS, pp. 183–196. doi: 10.1007/978-981-16-7334-4_14.

[18] 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.

[19] 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.

[20] 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.

[21] Liu, F. T., Ting, K. M. and Zhou, Z. H. (2012) 'Isolation-based anomaly detection', ACM Transactions on Knowledge Discovery from Data, 6(1). doi: 10.1145/2133360.2133363.

[22] 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.

[23] Pollet, T. V. and van der Meij, L. (2017) 'To Remove or not to Remove: the Impact of Outlier Handling on Significance Testing in Testosterone Data', Adaptive Human Behavior and Physiology, 3(1), pp. 43–60. doi: 10.1007/s40750-016-0050-z.

[24] Rousseeuw, P. J. and Hubert, M. (2018) 'Anomaly detection by robust statistics', Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(2), pp. 1–14. doi: 10.1002/widm.1236.

[25] 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.

[26] 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.

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

[28] 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.

[29] 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.

[30] 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.

[31] 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.

PROFILES

KONG YIH HERN is a Master student, studies in Tunku Abdul Rahman University College, Kuala Lumpur. He has a Diploma and Bachelor's degree in Computer Science and Software Engineering. His specialization is in forecasting and predictive analysis, interest in utilizing forecast tools as a solution to address business problems.
Email address: kongyh-wa15@student.tarc.edu.my

LIM KHAI YIN received her Master of Science (Artificial Intelligence) and PhD (Visual Analytics) in 2009 and 2017, respectively. Currently, she is a senior lecturer in TAR UC. Her research interests include computer vision, medical image processing, machine learning, and data analytics. She has also published papers both in journal and conferences relating to artificial intelligence (AI) and computer vision, specifically in fuzzy theory, neural networks, and medical imaging during her postgraduate study. She has experiences in guiding final year project (FYP) students on the topics that involves AI and computer vision. Besides that, she has been involving in some industry projects, mainly in the topics of classifying painting styles and engagement prediction in social media.
Email address: limky@tarc.edu.my

CHIN WAN YOKE received B.Sc. and Ph.D. degrees from Universiti Teknologi Malaysia in the year 2009 and 2016, respectively. She is currently a postgraduate programme leader at Tunku Abdul Rahman University College and is actively involved in the activities organized by the Faculty of Computing and Information Technology, Center for Data Science and Analytics, and Centre for Postgraduate Studies and Research. Her main research focus is mathematics and statistics. She is a multidisciplinary researcher who is involved in several projects including data analytics, predictive modeling, optimization, statistical hydrology, painting styles classification, and social media engagement.
Email address: chinwy@tarc.edu.my