

Article

Social media marketing in the sales volume prediction for the Lolita fashion brand

Ton Chaing^{1*}, Hsin Rau¹, Wei Jung Shiang¹ and Jon Luen Chiang²

¹ Department of Industrial and Systems Engineering, Chung Yuan Christian University, Taoyuan City 320314, Taiwan, ROC; tchiang.0514@gmail.com (T.C.); hsinrau@cycu.edu.tw (H.R.); wjs001@cycu.edu.tw (W.J.S.)

² Department of Electrical & Electronic Engineering, City, University of London, London EC1V 0HB, United Kingdom; jlarchie@gmail.com (J.L.C.)

* Correspondence: tchiang.0514@gmail.com (T.C.); Tel.: +886933089188, +886-2-2450-8153

Abstract: Despite extensively investigating the impact of social media on fashion products' marketing, little evidence is available on how the platforms influence sales prediction. Focusing on Lolita fashion, this study investigates the impact of social media marketing on the sales volume prediction of fashion products. Essentially, we analyzed marketing data, including comments, likes, and shares from the Weibo social platform, to forecast future sales, examine how to enhance profit performance, and make production decisions. Using a quantitative approach, we tested three different prediction models, including multiple regression, decision tree, and XGBoost. The results revealed that increasing comments and decreasing the number of likes could significantly improve the sales volumes of Lolita products. In contrast, shares exerted a less significant impact on sales. Regarding prediction models, XGBoost was found to be the best method. In the fashion industry, social media is a useful tool for forecasting market trend. A limitation of this study is that only one social media platform was used to extract data, which might limit the generalization of the findings.

Keywords: Lolita fashion; multiple regression; decision tree; social media; XGBoost

1. Introduction

Lolita fashion is a dressing culture portraying young women dressed like frilly Victorian dolls. The fashion is famous in Japan, and the movement is growing in America via the Live Journal with over 8000 members in its Elegant Gothic Lolita community. Over time, Lolita fashion has evolved but is still considered a celebration of cuteness, modesty, beauty, and femininity. The basic elements of a Lolita outfit comprise a bell-shaped skirt supported by a petticoat and worn along with bloomers [1]. Other accessories comprise over-the-knee socks, hair bow, headdress, or beret worn with round-toe shoes [1]. The high-priced nature of the fashion is mostly attributable to the high-quality laces and fabric that go together with the clothing [2]. Internet marketing management can be used to market Lolita products through various strategies and options available for predicting sales volumes. The cost–volume–profit analysis can be performed to minimize resource wastage and maximize sales volume outputs [3]. Using the share, comment, and like data obtained from the marketing pages available on Facebook, Weibo, and Instagram, the expected sales can be predicted, thereby determining the economic order for sales planning. In addition, sales forecasts can be a predictor of how much revenues can be anticipated in a given period, thereby determine how much to manufacture and, thus, minimize production costs. Moreover, effective management of inventory can be realized through effective sales forecasting. Social media data are effective in gathering buyers' opinions, including their past experiences and current expectations about products and services. By incorporating these views and expectations into products and services, significant forecasts of sales volumes can be made [4]. By applying social media data, enterprises can

predict sales volume and plan production, thereby maximizing resource use. Furthermore, reasonable and effective inventory management can then be attained. Maximization of productions decreases the overstock of products, thereby achieving lean productions [5].

1.1 sales volume prediction statement

The sales volume prediction is a crucial aspect of business profit maximization that can enhance brand marketing management. A wide array of traditional time-series models used to predict future product sales have relied on previous sales data, which are historical and seasonal [6]. The models often overlook the dynamic and crucial impact that recent events have in critically influencing sales. Online review efficiency in promoting sales and forecasting futures sales volume can markedly assist in brand marketing management [7]. Moreover, the traditional time-series models fail to gather the prevailing sentiments and opinions about products and companies, which are crucial in forecasting sales patterns [8]. Social media platforms provide data on consumers' opinions and likes, which can be applied in models to comprehend how to increase the sales volume of a product. Owing to the restricted local demand for Lolita fashion brands, investigating how to effectively use social media to garner attention and generate consumer awareness is crucial to increase sales. This study aims to investigate social media marketing in the sales volume prediction with a particular focus on Lolita brand marketing management. Thus, the study objectives include the following:

- (1) To illustrate how social media marketing can be used to forecast future sales.
- (2) To investigate how sequential prediction of sales can enhance profit performance.
- (3) To investigate how enterprises can determine productions based on sales and arrange production according to sales forecast data to evade product overstocking.

1.2 Research significance

An understanding of these trends, if established, will assist the marketing plans and strategies adopted in future sales management and resource use. Besides, examining Lolita brand marketing in this study will explain how Lolita can be popularized using online platforms. As Lolita is not a mainstream fashion brand, understanding how to effectively market it using social media will ensure that both domestic and foreign consumers can be targeted to increase sales. Moreover, the findings will enable retail shops of Lolita products to understand how they should engage with consumers on social media to increase sales. In addition, our findings will address the periodic forecasting problems noted through the use of other models. In particular, social media data use will be presented as a supplement to the existing analytical methods available for sales volume forecasts. By establishing an understanding of the model application, effective inventory management will also be understood by marketing managers to optimize resource use, thereby maximizing sales revenues. Furthermore, the frequency patterns for sales will be established, thereby improving forecasts for future sales. Finally, the enthusiasm of the sales staff will be mobilized through the establishment of a more reliable method of sales forecasts.

2. Methodology

2.1 Data collection

The data were collected on social media platform Weibo by Lolita brand of Long Ears and Sharp Ears studio, and Pinkup Lolita (<https://lnears.taobao.com>, <https://pinkuplolita.taobao.com>) authorized calculations from 2014 to 2020. During data collection, owing to the influence of COVID-19, consumers' behavior changed from offline shopping to online shopping. Thus, the accuracy of data statistics has become even more helpful to the brand's sales forecast. In the search process, the main focus was on different Lolita products in which the researcher identified each brand and recorded the number of likes, shares, and comments in an Excel file. We considered 36 different Lolita skirts, such as

Machinacium, Kingdom, Deer Music, stardust, Gold Fish, and Hymn of Love. After extracting data from social media, the researcher sought corresponding data of the sales volume for the Lolita brands from dealers to determine whether a higher number of likes, shares, and comments translated to higher sales.

2.2 Data analysis

From the collected data, the dependent variable was sales volume, while the independent variables were social media parameters of like, comment, and share. The data were analyzed using different techniques, including correlation analysis in R language, multiple regression analysis, forecast using a decision tree, and analysis using XGBoost. Each data analysis approach has been summarized in the next sections.

2.2.1 Correlation analysis in R language

The first data analysis procedure involved a correlation analysis to determine the correlations among the four variables, including the number of sales, likes, shares, and comments. The data of the four variables were first imported into the R-language program, followed by importing the associated package, including ggthemes and ggcorrplot, which were used to ascertain the correlation coefficient for the variables. Besides calculating the correlation matrix, the R-program visually presented the data. When interpreting the data, the main focus was to identify correlation coefficient values > 0.7 , which suggests a high correlation between two variables. Section 3 presents the correlation output from the R-language.

2.2.2 Multiple regression analysis

A multi-regression analysis was performed in the R program to determine the regression model, which connected sales volume with the independent variables of likes, shares, and comments. In addition, the multiple regression analysis aimed to help separate the significant social media factors influencing sales from the less significant ones by comparing probability, P values with the α value of 0.05. The prediction equation obtained from the analysis was decisive in ensuring that future sales of Lolita fashion products can be estimated per social media activity including likes, shares, and comments.

2.2.3 Decision tree

A decision tree was developed in the R program to assist in predicting the sales volumes for given values of comments, likes, and shares. The objective was to determine whether using a decision tree in forecasting the sales volumes of Lolita brands shows higher accuracy than using multiple regression analysis. In the process, a decision tree was developed, and a corresponding graph of predicted and original data was indicated.

2.2.4 XGBoost

XGBoost is an algorithm that uses a decision tree model and is considered highly accurate in solving regression equations and prediction problems compared with the existing artificial neural networks. We used XGBoost to ensure minimum root mean square error when fitting the predicted model over the real values. The main advantages of XGBoost modeling include regularization that prevents overfitting. In addition, XGBoost uses both the first and second derivatives to ensure minimum errors in prediction. Finally, the granularity feature of XGBoost ensures that modeling incorporates parallel optimization, which considers missing data and ensures efficient and optimized output.

3. Result Analysis

3.1 Correlation analysis

A correlation analysis was conducted in R language software to identify the correlation between different research variables, including share, comment, like, and sales volume. Figure 1 shows the visualization of the data correlation matrix.

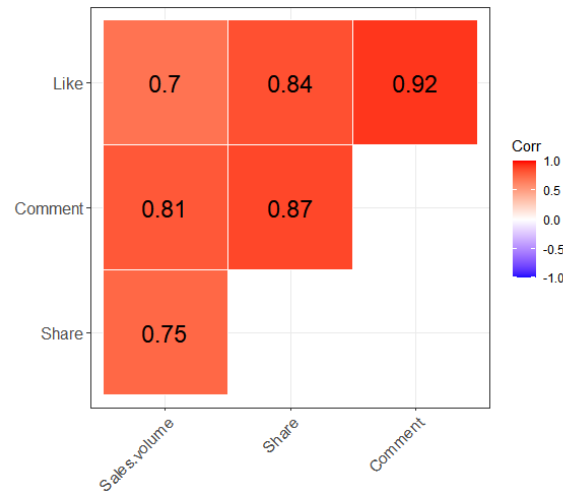


Figure 1. Correlation of variables visualized in the matrix.

Figure 1 shows that the correlation coefficients of all the pairs of variables were >0.7 . The high positive correlation between the variables indicated that sales volume positively correlated with shares, likes, and comments. Another interesting result was that sales volume had the highest correlation with comments, implying that customers who commented about the quality or characteristics of a product were likely to purchase compared with those who shared or liked social media posts on social media. To gain more insights into the significance of the correlations shown in Fig. 1, P values of the correlations were determined; Table 1 summarizes the results.

Table 1. P values of the correlation between variables

	Shares	Comments	Likes	Sales volume
Share	0e+00	0	6.00000e-15	1.00000e-15
Comment	0e+00	0	0.00000e+00	0.00000e+00
Like	6e-15	0	0.00000e+00	5.09657e-10
Sales volume	1e-15	0	5.9657e-10	5000.00000e+0

Table 1 shows that sales volume significantly correlated with share ($P = 1.00e-15$), comment ($P = 0.00e+00$), and like ($P = 5.09e-10$), as all P values are <0.05 . These findings suggested that encouraging customer interaction on the social media platform Weibo through likes, shares, and comments could significantly affect sales volume. The visualization of sales volume correlation with comments, likes, and shares was performed by creating scatter plots and fitting lines, as shown in Fig. 2(a–c).

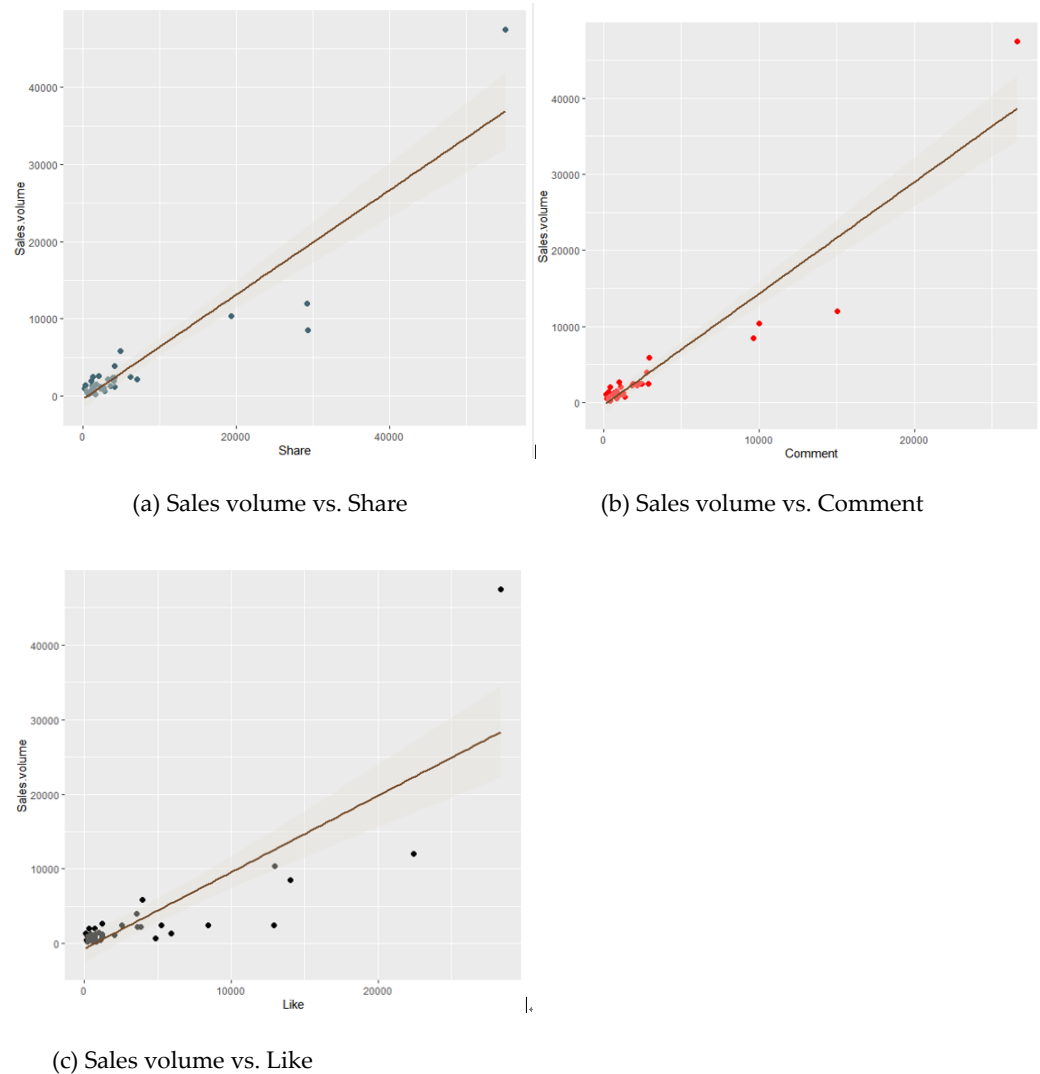


Figure 2. Scatter plot and fitting line of sales volume vs. (a) Share, (b) Comment, and (c) Like.

Each scatter plot in Fig. 2 demonstrated that the fitting line had a positive upward inclination, suggesting that increasing each factor loading correspondingly increased the sales volume. Although conducting correlation analysis was crucial to elucidate the correlation between individual social media variables with sales volume, even more essential was exploring the joint impact of the three social media variables on sales volume. Thus, we performed a multiple regression analysis, the results of which are indicated in the next section.

3.2 Multiple regression analysis

The multiple linear regression aimed to identify how a particular dependent variable changed when the different independent variables were changed. In this study, multiple regression output from R-language software was obtained, as shown in Fig. 3.

```

> #3
> model <- lm(Sales.volume~Share+Comment+Like,data = data)
> summary(model)

Call:
lm(formula = Sales.volume ~ Share + Comment + Like, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-7918.0 -427.6  -42.4   486.7  6103.5

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  341.7815   458.9762   0.745 0.461911
Share        -0.2660    0.1979  -1.344 0.188302
Comment       2.7869    0.4881   5.710 2.52e-06 ***
Like         -0.6504    0.1703  -3.820 0.000579 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2241 on 32 degrees of freedom
Multiple R-squared:  0.9287,    Adjusted R-squared:  0.922
F-statistic: 138.9 on 3 and 32 DF,  p-value: < 2.2e-16

```

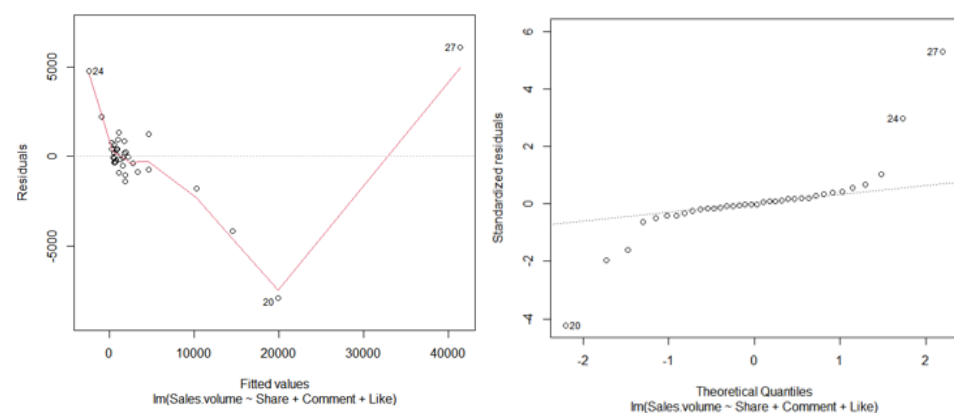
Figure 3. Multiple regression output.

As shown in Fig. 3, the regression coefficients were used to develop the regression model that could be used to predict the sales volume based on shares, likes, and comments. The regression equation can be summarized as follows:

$$\text{Sales volume} = 341.78 - 0.266 * \text{Share} + 2.787 * \text{Comment} - 0.65 * \text{Like} \quad (1)$$

Equation (1) reveals that when the joint impact of the three social media variables on sales variable is considered, shares and likes have a negative correlation of coefficient (r) = -0.266 and -0.65 , respectively. The results indicated that increasing the number of shares and increasing the number of likes could lead to a reduction in the sales volume. In contrast, sales volume positively correlated with comments ($r = 2.787$), suggesting that increasing the number of comments about Lolita skirts led to higher sales. The regression t -test in Table 2 revealed that in the regression model, comment ($P = 2.52e-06$) and like ($P = 0.000579$) variables significantly affected sales volumes owing to $P < 0.05$, while shares exerted a less significant impact ($P = 0.188$). Another key finding from Table 2 is the R^2 value of 0.929, which shows that the regression model developed from the analysis has a high accuracy of 92.9% in predicting future sales volume if social media variables are provided, including likes, shares, and comments.

In conducting multiple regression, the other crucial analysis performed was a normality test to ascertain the assumption made during regression that the residual errors are approximately 0. Usually, for normality tests, the normal Q-Q plots produce linear relationships. Figure 4(a and b) summarize the residual plots and normal Q-Q plots.



(a) Residual plots

(b) Normal Q-Q plots

Figure 4. (a) Residual plots of multiple regression and (b) normal Q-Q plots.

In Fig. 4(a), the residual plot indicates that residual errors are close to 0, with a few outliers, indicating a good fitting effect and that the data used were normally distributed. The assumption of normality is confirmed in Fig. 4(b), showing that most of the scattered plots fit within a straight line. Thus, the multiple regression model obtained in the analysis is valid, as there were no inherent errors in the dataset gathered.

3.3 Segmentation training set and test set

The multiple regression model developed was applied using a dataset to identify its accuracy by considering a fitting diagram. The first step involved developing a segmentation training set and a test set in R-language software, as summarized in Fig. 5. In the segmentation, 30% of datasets were test sets (11 samples), while 70% were training sets (25 samples).

```
#4
sub <- sample(nrow(data),nrow(data)*0.7)
train_data <- data[sub,]
test_data <- data[-sub,]
```

test_data	11 obs. of 4 variables
train_data	25 obs. of 4 variables

Figure 5. Training and test datasets.

3.4 Multiple regression prediction

We used the regression model to predict sales volumes based on the original data of likes, shares, and comments. In addition, we aimed to check the disparity between the fitting graph and plots of the original values of sales volume. The main parameter assessed was the root mean square error (RMSE). We developed the R-language algorithm and output for multiple regression prediction as shown below:

```
R> #5
R> model_f1 <- lm(Sales.volume~Share+Comment+Like,data = train_data)
R> test_data$predict1 <- predict(model_f1,test_data)
R> RMSE1 = (sum((test_data$predict1-test_data$Sales.volume)**2)/nrow(test_data))**0.5
R> RMSE1
[1] 2931.916
```

The fitting graph of the predicted values and plots for original values were obtained as shown in Fig. 6. Based on the results and algorithm run in Fig. 6, we obtained the RMSE of the fitting plot as 2931.916. Essentially, RMSE indicates the level within which a model can accurately predict a specific response variable; hence, a lower RMSE value is often preferred. The high RMSE value of 2931.916 suggested that the regression model obtained had some limitations in prediction. The finding is confirmed in Fig. 6, which shows that in the sample with ID 8, a significantly large difference exists between the real and predicted value. Nevertheless, the other real values in Fig. 6 are relatively consistent with the predicted values, suggesting that despite certain limitations, the regression model developed is fairly accurate and can be used for predicting sales volume.

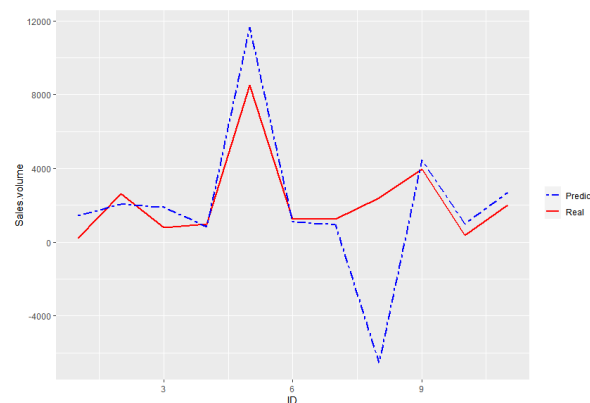


Figure 6. Comparison between real sales volume data and the predicted values using the regression model.

In addition, Fig. 6 shows that the sample with ID 5 has a large difference between the peak of its real and predicted sales volume value. The large difference reinforces the idea that a more accurate model is needed. Hence, we further explored decision tree and XGBoost regression models.

3.5 Forecasting with decision tree

The underlying idea of decision trees in forecasting is that an algorithm is developed to help in automatically dividing a large dataset into different categories. The assumption made in the technique is that an original dataset gathered is unstructured and difficult to understand. After dividing the dataset into classes using an algorithm, multivariate models can be used to predict the values for classes of interest. In this study, the algorithm developed in R-language software to assist in decision tree forecasting, as well as the resultant RMSE value, are shown below:

```
R> tree.fit <- rpart(Sales.volume~Share+Comment+Like, data= train_data )
R> par(mfrow=c(1,1))
R> test_data$predict2 <- predict(tree.fit, test_data)
R> RMSE2 = (sum((test_data$predict2-test_data$Sales.volume)**2)/nrow(test_data))*0.5
R> RMSE2
[1] 3900.498
```

Figure 7 shows the decision tree developed using the above-described algorithm.

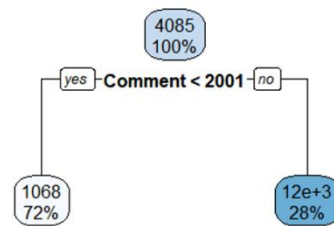


Figure 7. Decision tree used for forecasting.

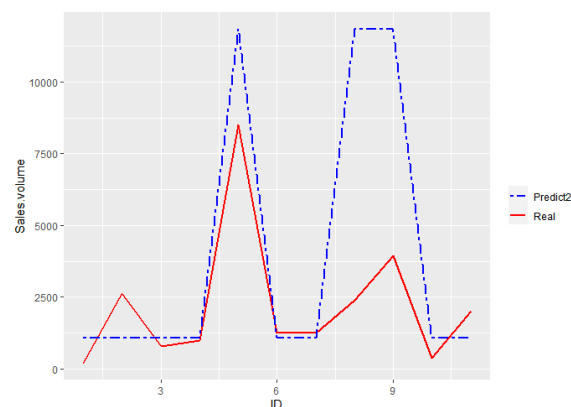


Figure 8. Comparison between real sales volume data and the predicted values using the decision tree method.

From Figure 8 and RMSE value obtained, certain key points are observed. First, the obtained RMSE value of 3900.498 for the decision tree prediction model was comparatively higher than the 2931.916 value of the regression model. In addition, the fitting graph shown in Fig. 8 exhibited larger differences from the actual values, suggesting that the decision tree model was less accurate compared with the multiple regression model. Spe-

cifically, Fig. 8 shows that in the sample with ID 9, the real sales volume of 3750 is significantly lower than the predicted value of over 11250. These results emphasized the need to further explore XGBoost to determine whether it can provide a more accurate regression prediction model.

3.6 XGBoost

The use of XGBoost algorithm is a recent trend in regression and considered the best optimization technique about data prediction. Although XGBoost uses decision tree principles, it is designed to use a gradient boosting approach, which optimizes operations through decision tree splitting, parallelization where outer and inner loops are executed concurrently to increase execution time, and the algorithm also optimizes hardware by considering the available disk space in its execution. The other benefit of XGBoost is that it uses regularization to prevent overfitting. Essentially, regularization involves shrinking the coefficients of less significant factors to zero, as well as eliminating specific features that exert an insignificant impact. The strategy prevents the development of erroneous models and ensures that only key features in regression are used in developing a fitting model. Thus, it is expected that XGBoost produces superior output than decision tree forecast method. In this study, the algorithm used in XGBoost regression prediction, as well as the corresponding RMSE, is summarized below:

```
> traindata1 <- data.matrix(train_data[,c(1:3)])
> traindata2 <- Matrix(traindata1, sparse=T)
> library(xgboost)
> traindata1 <- data.matrix(train_data[,c(1:3)])
> traindata2 <- Matrix(traindata1, sparse=T)
> traindata3 <- train_data[,4]
> traindata4 <- list(data=traindata2, label=traindata3)
> dtrain <- xgb.DMatrix(data = traindata4$data, label = traindata4$label)
> testset1 <- data.matrix(test_data[,c(1:3)])
> testset2 <- Matrix(testset1, sparse=T)
> testset3 <- test_data[,4]
> testset4 <- list(data=testset2, label=testset3)
> dtest <- xgb.DMatrix(data = testset4$data, label = testset4$label)
> xgb <- xgboost(data = dtrain, max_depth=6, eta=0.5, objective='reg:linear', nround=25)
[12:02:02]WARNING:amalgamation/./src/objective/regression_obj.cu:174:reg:linear is now
Deprecated in favor of reg:squarederror.
[1] train-rmse:7525.872559
[2] train-rmse:5569.117188
[3] train-rmse:4148.380371
[4] train-rmse:3095.326904
[5] train-rmse:2314.129639
[6] train-rmse:1735.6288477
[7] train-rmse:1301.855713
[8] train-rmse:976.688477
[9] train-rmse:732.294861
[10] train-rmse:549.496460
[11] train-rmse:412.433960
[12] train-rmse:309.706604
[13] train-rmse:232.635849
[14] train-rmse:174.796066
[15] train-rmse:131.458694
[16] train-rmse:98.821480
[17] train-rmse:74.385635
[18] train-rmse:56.124573
[19] train-rmse:42.488651
[20] train-rmse:32.337677
[21] train-rmse:24.293238
[22] train-rmse:18.336798
[23] train-rmse:13.848682
[24] train-rmse:10.365655
[25] train-rmse:7.773936
R> test_data$predict3 = round(predict(xgb,newdata = dtest))
R> RMSE3 = (sum((test_data$predict3-test_data$Sales.volume)**2)/nrow(test_data))**0.5
R> RMSE3
[1] 1290.707
```

The XGBoost algorithm analysis reveals that RMSE was 1290.707, which was lower than that of the multiple regression model (2931.916), as well as that of the decision tree model (3900.498). These results suggested that using XGBoost model in the prediction of

sales volume produced the smallest error and confirmed the efficacy of XGBoost regression as the best prediction model. In addition, the result of RMSE for XGBoost was confirmed in the fitting plot of real and predicted values shown in Fig. 9.

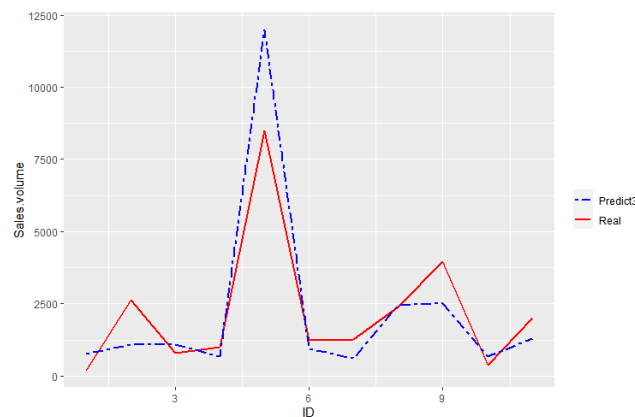


Figure 9. Comparison between real sales volume data and the predicted values using the XGBoost regression method.

Figure 9 shows that the peaks of the real and predicted graphs are more aligned with each other than the previous two graphs developed of the multiple regression model and decision tree model. Thus, Fig. 9 confirms that XGBoost is the best prediction model to use in regression.

4. Discussion

4.1 Usage of social media marketing to improve sales of fashion products

One of the main objectives of this study was to investigate the efficacy of using social media to increase sales of fashion products. We first noted that the regression model connecting social media factors and sales volume was statistically significant, suggesting that using social media to market fashion products could lead to a significant change in sales volume. This result corroborated a previous study [9], which reported that engagement with consumers on social media platforms facilitated customer service and allowed businesses to address consumers concerns, as well as provided accurate information regarding their products. The net effect is that consumers can avoid manipulation by middlemen and unscrupulous retailers who overcharge fashion products or sell counterfeit products. Regarding our study, the result implied that luxury fashion brands, such as Lolita, should intensify their social media marketing efforts to attain higher sales. This finding aligned with a previous study [10], which demonstrated that posting products on social media increased the visibility of such brands as consumers liked, commented, and shared the posts. Thus, more than engaging with consumers, luxury fashion brands should strive to increase their visibility on social media by gaining more followers and involving social media influencers.

In addition, this study analyzed how different social media variables, including likes, shares, and comments, influence sales volumes of fashion products. The regression analysis revealed that comments and likes exerted a significant impact on sales volumes, while shares exerted a less significant impact; this result implied that when using social media, Lolita fashion brand should prioritize strategies that influence likes and comments by customers. The obtained results revealed that comments positively correlate with the sales volume of Lolita products, which is attributable to customers providing their experiences with the Lolita fashion brand, thereby influencing other social media users to purchase such products. Moreover, comments ensure that Lolita brand managers can clarify customers concerns, as well as provide information about Lolita products, creating awareness and improving the confidence of consumers regarding their purchase decisions. This finding aligns with a previous study [11], which highlighted that when brands engage with

consumers on social media, they can identify areas of product improvement, as well as the most popular products, where they should focus their production and marketing. In our study, Lolita's usage of social media enabled them to determine the popular brands among consumers by checking at comments or the images posted, influencing them to determine the best strategy of maximizing sales. The obtained result corroborated a previous study [12], which reported that comments on social media enabled brands to keep track of information so that they do not need to conduct surveys to gain insights into what customers perceive about a particular product. Thus, considering customers' comments on social media enabled Lolita brand managers to understand the emerging trends in the fashion sector over a specific period, thereby enabling them to adjust to the market needs.

Although the multiple regression analysis demonstrated that social media likes exerted a significant impact on sales, it was surprising to realize a negative correlation between likes and sales. Essentially, the finding implied that increasing the number of likes for Lolita brand posts led to a reduction in sales volume; however, this is attributable to the luxury brand of Lolita products. The implication is that its consumers delight in the brand being exclusive for a select few people who are considered special because they are perceived as a high social class. Consequently, engaging a large number of people on social media on the Lolita luxury brand can contribute to the brand losing its exclusivity; hence, previous customers might stop buying the brand. In this respect, for luxury brands, having many likes might not be a positive trend and, instead, might discourage potential consumers from purchasing the marketed product. Thus, for luxury fashion brands, such as Lolita, social media managers should avoid engaging influencers who can lead to a significant increase in the number of likes for posts and instead engage with customers directly.

The other finding of the multiple regression analysis was that social media shares had a less significant and negative relationship with sales volumes; this was unanticipated because sharing posts on social media is expected to increase visibility and ensure friends and family can as well view a recommended product and even make the purchase. However, the finding implied that when seeking to increase the sales volumes for luxury fashion brands, the number of social media shares should not be a consideration. Thus, sharing increases the visibility of products being marketed on social media. Nevertheless, a negative correlation between social media shares and visibility was attributed to shares increasing the product popularity, thereby causing a loss of exclusivity upon which the brand is centered.

Moreover, the regression model revealed that the comments positively correlated with sales volume such that an increase in comments was accompanied by an increase in the quantities sold of Lolita fashion products. Meanwhile, the multiple regression suggested a statistically significant correlation between comments and sales volume. These findings corroborated a previous study [13], which highlighted that consumers utilized social media data to inform their perception of fashion products. Moreover, the authors added that consumer demand was aligned per the type of information mined from social media, leading to an increase in sales volume. Meanwhile, another study [14] highlighted the significance of customer engagement in the comment section in improving sales; the authors explained that consumer loyalty improved when users felt that their views were regarded. The implication of the findings is that social media comments do not generate sales on their own without the intervention of online marketers. Essentially, business response to comments plays a vital role in winning the trust and confidence of consumers for repeated purchase trends. Hence, Lolita fashion benefited from the continued efforts of experienced digital marketers who interacted with consumers on social media and forwarded important views to the management for action.

4.2 Accuracy of sales prediction models & implications

This study also sought to investigate the best model among multiple regression, decision tree, and XGBoost, which can be used to precisely predict the sales volumes of Lo-

lita fashion. For the multiple regression model, the results demonstrated that the prediction accuracy was up to 92.9%, and the outcome relied on the availability of social media data such as likes, shares, and comments. The implication of the findings is that the accuracy of the multiple regression model in predicting the sales volume for both present and future relies heavily on the availability of large data in the form of dependent and independent variables. Thus, the model is applicable to the apparel industry, whereby data from the social media marketing campaign is present. A previous study has discussed the accuracy of the multiple regression method in the presence of numerous data, providing a solution to the limitations of the baseline sales prediction models [15]. The authors reported that incidences of many variables rendered baseline methods less accurate owing to the increased complexities of the analysis process. However, in multiple regression, the complexities are taken care of but with assumptions on the residual errors, which are always taken to be zero. Based on our results, the errors decreased the accuracy value; thus, the model was considered relatively accurate. These findings aligned with another study [16], which noted that inaccuracy in predicting sales volume in the fashion industry was attributed to high levels of uncertainties such as that of customer preferences. Alternatively, the assumptions made in considering social media data contributed to erroneous conclusions about the models' efficiency. Based on the revelation on the accuracy of the multiple regression model, we found that although the method is superior in the sales volume predictions, the assumption aspect presents limitations, which merits a further consideration of improve models. Nonetheless, the model is suitable for initial predictions in the fashion industry, as the accuracy is adequately high for a convincing conclusion.

Compared with the decision tree model, the results revealed that the multiple regression method was more accurate in predicting sales volumes in the fashion industry. Principally, the RMSE value was higher for the decision tree model than for multiple regression. The findings aligned with a previous study [17], which demonstrated noted that the efficiency of various sales prediction techniques depended on the field of application. In that regard, the prediction tree model for predicting sales volume in the fashion industry encountered several hurdles, which negatively affected the accuracy values. According to another study [18], the textile-apparel industry to which Lolita fashion belongs exhibits dynamics that calls for more intelligent models of prediction. Thus, the decision tree prediction model only caters for the basic data and becomes inaccurate when complex variables are included. With the increasing use of social media platforms in promoting Lolita fashion products, it is evident that relying on decision trees for future and present the sales volume prediction can lead the business to wrong decisions. While the model may only function with simple variables, it becomes unfit for use in the fashion industry, whereby a study [19] noted that the higher the visibility in social media, the higher the data and, hence, enhanced accuracy of sales data prediction. In essence, the precision of the prediction is defined by the availability of a broad range of social media data. Consequently, the best prediction model is one that can effectively handle the variables with minimum complications and produce the least RMSE. Hence, the multiple regression model is better than the prediction tree in terms of data handling and accuracy; this result implies that some fashion businesses that have relied on decision trees in predicting sales volume using social media data have been exposed to erroneous figures. Such values result in poor allocation of resources and production scheduling, which contributes to either overstocking or understocking the market.

Meanwhile, an evaluation of XGBoost revealed it to be the best of the three models. The model used the gradient boosting approach to optimize operations and, thus, enhanced the execution time of the variables. Regarding the RMSE, XGBoost produced the least value such that the graph of real and predicted values aligned more closely than those of multiple regression and decision tree models. Thus, the XGBoost model is one of the most intelligent software available that can simplify the complexities. Hence, the model is suitable in planning for stocking and target markets in the future of Lolita fashion. Using the values obtained from social media, the findings imply that the prediction method used is a key indicator of the accuracy level of the sales prediction. For Lolita

fashion future sales prediction, the most appropriate approach would be the XGBoost model, followed by the multiple regression model. As noted previously [20], prediction accuracy also depends on the social media platform used, with the common ones being Twitter, Facebook, and Google trends. Thus, the accuracy of future sales prediction models considers the source and type of data utilized. Consequently, this study implies that each fashion line settles for a particular prediction model based on the level of social media exposure and the type of data mined for analysis. In that respect, even the decision tree model can be useful in some circumstances that involve simple data of social media activities. The sentiment reflects the argument made by another study [21] that the use of low-frequency data is suitable in ensuring the computation is relatively easy results and that the data accuracy used in predicting sales is improved. Fundamentally, an accurate prediction of sales leads to a certainty in production and higher profits owing to decreased wastage. For Lolita fashion, the most reliable sales forecast method is one that will incorporate all the data from social media and present the most accurate prediction of consumers' preferences. Based on the ability to identify the future characteristics of the market for a particular brand, a prediction model is considered efficient. Accordingly, the XGBoost model is the best that should be applied for future sales prediction.

5. Conclusions and Recommendations

5.1 Conclusions

This study shows that predicting future sales is possible by analyzing the user data from social media platforms. By having a Lolita fashion presence on various social media platforms, significant related data, including consumer likes, shares, and comments, can be mined and used for future sales volume prediction. However, the accuracy of predictions depends on the prediction model used. Based on the results, the XGBoost model is the superior sales volume prediction technique for the Lolita brand, as it has the lowest RMSE compared with multiple regression and decision tree models. Hence, it can be concluded that social media provide relevant data useful in the sales volume prediction in the fashion industry, with XGBoost as the best model.

In addition, profit performance improvement directly correlates with improvement in sales of fashion products. The regression model reveals a statistically significant correlation between social media factors and sales volume. Specifically, social media variables, such as likes, share, and comments, on the fashion brands determine the trend of sales. The most influential variable is likes, while the least is shares. By considering the impact of likes and shares on sales volume and pattern, this study establishes that social media reactions to a fashion post influence the attitude and judgment of consumers.

Concerning the Lolita brand, likes and shares negatively correlate with sales volume. Essentially, more likes and sharing increase the brand visibility on social media, making it popular and, hence, common and less attractive to consumers. Hence, the use of social media negatively affects the sales volume of fashion products, leading to a reduction in profits. As such, care must be taken in utilizing the platforms to minimize incidences and adverse effects of brand familiarity. Furthermore, this study demonstrates that the level of production depends on the sales volume. Thus, a highly accurate sales prediction model is the key to determining the number of goods to stock. In other words, consideration of consumer reactions on social media is helpful in foretelling the likely quantity to be purchased on a specific fashion brand. Regarding Lolita fashion, numerous likes and shares guide the business to limit stock, as it indicates a decline in the sales volume in the future as the product becomes common. Hence, social media is a useful tool for the fashion industry to project market trends in terms of consumer shopping behavior, sales, production, and profit levels of an enterprise.

5.2 Recommendations for future research development and practice

The concept of social media use in sales enhancement and prediction in the fashion industry is crucial, as most enterprises are embracing the platforms for marketing campaigns and general visibility. To elucidate the social media influence, future studies must involve the actual assessment of consumer experience with the frequent fashion brand posts. The outcome of such studies will support the statistical claims presented in this study. Moreover, engaging consumers will explain the social media user mindset regarding fashion brands postings. Consequently, respective managers will be in a position to effectively strategize social media marketing approaches. In addition, future studies should consider involving the qualitative analysis of testimonies from management and social media marketers of fashion products such as Lolita representatives. The views of enterprises would add value to the sales trend observed in this study and offer explanations for the choice of sales prediction models.

Regarding practice, this study establishes that social media data miners and analysts for fashion enterprises should take great caution as to the type of data utilized in sales determination. Notably, the trend of likes and shares for fashion products correlates inversely with sales compared with other product industries. Thus, fashion marketers must effectively strategize online marketing such that the brand does not become common yet remains the most popular in terms of quality and pricing. Meanwhile, a choice of the most appropriate sales prediction model is key to the future performance and sustainability of fashion enterprises. Finally, data analytics should choose a model with the least RMSE, such as XGBoost.

Author Contributions: conceptualization, T.C. and H.R.; methodology, T.C. and H.R.; software, T.C. and J.L.C.; validation, T.C., H.R. and W.J.S.; formal analysis, T.C. and J.L.C.; investigation, T.C.; resources, T.C. and H.R.; data curation, T.C. and W.J.S.; writing—original draft preparation, T.C.; writing—review and editing, H.R. and W.J.S.; visualization, T.C.; supervision, T.C. and H.R.; project administration, T.C. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data that support the findings of this study are available from the first or corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Haijima, A. Japanese popular culture in Latvia: Lolita and Mori fashion. *Regional studies*, **2013**, pp. 31-56.
- Kang, Z.Y.; Cassidy, T. Lolita fashion: a trans-global subculture. *Fashion, Style & Popular Culture*, **2015**, pp. 371-384.
- Kirlioğlu, H.; Baral, G. In the uncertainty conditions cost-volume-profit analysis which is used fuzzy logic. In *International Symposium on Sustainable Development*. **2012**.
- He, W.; Guo, L.; Shen, J.; Akula, V. Social media-based forecasting: a case study of tweets and stock prices in the financial services industry. *Journal of Organizational and End User Computing (JOEUC)*, **2016**, 28(2), pp. 74-91.
- Shi, Y.; Karniouchina, E.; Uslay, C. (When) can social media buzz data replace traditional surveys for sales forecasting? *Rutgers Business Review*, **2020**, 5(1), pp. 43-60.
- Ahn, H.I.; Spangler, W.S. Sales prediction with social media analysis. In *2014 Annual SRII Global Conference*. IEEE. **2014**, pp. 213-222.
- Ngo-Ye, T.L.; Sinha, A.P. Analyzing online review helpfulness using a regressional relief F-enhanced text mining method. *ACM Transactions on Management Information Systems (TMIS)*, **2012**, 3(2), pp. 1-20.
- Liu, B. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, **2012**, 5(1), pp. 1-167.
- Hanna, R.; Rohm, A.; Crittenden, V. We're all connected: the power of the social media ecosystem. *Business Horizons*, **2011**, 54(3), pp. 265-273.
- Popp, B.; Woratschek, H. Introducing branded communities in sport for building strong brand relations in social media. *Sport Manage. Rev*, **2016**, 19(2), pp. 183-197.
- Dholakia, U.M.; Bagozzi, R.P.; Pearo, L.K. A social influence model of consumer participation in network-and small-group-based virtual communities. *International Journal of Research in Marketing*, **2004**, 21(3), pp. 241-263.
- Curran, K.; Graham, S.; Temple, C. Advertising on Facebook. *International Journal of E-Business Development*, **2011**, 1, pp. 26-33.
- Luo, X.; Zhang, J.; Duan, W. Social media and firm equity value. *Information Systems Research*, **2013**, 24(1), pp. 146-163.
- Barton, D.; Court, D. Making advanced analytics work for you. *Harvard Business Review*, **2012**, 90(10), pp. 78-83.
- Beheshti-Kashi, S.; Karimi, H.R.; Thoben, K.D.; Lütjen, M.; Teucke, M. A survey on retail sales forecasting and prediction in fashion markets. *Systems Science & Control Engineering*, **2015**, 3(1), pp. 154-161.
- Nenni, M.E.; Giustiniano, L.; Pirolo, L. Demand forecasting in the fashion industry: a review, *Int. J. Eng. Bus. Manage*, **2013**.

17. Thomassey, S. Sales forecasting in apparel and fashion industry: a review. *Intelligent fashion forecasting systems: models and applications*, **2014**, pp. 9-27.
18. Yu, Y.; Choi, T.M.; Hui, C.L. An intelligent fast sales forecasting model for fashion products. *Expert Systems with Applications*, **2011**, 38(6), pp. 7373-7379.
19. Giri, C.; Thomassey, S.; Zeng, X. The exploitation of social network data for forecasting garment sales. *International Journal of Computational Intelligence Systems*, **2019**, 12(2), pp. 1423-1435.
20. D'Avanzo, E.; Pilato, G.; Lytras, M. Using Twitter sentiment and emotions analysis of Google Trends for decisions making. *Program*. **2017**.
21. Wu, X.; Shi, B.; Dong, Y.; Huang, C.; Faust, L.; Chawla, N.V. Restful: resolution-aware forecasting of behavioral time series data. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, **2018**, pp. 1073-1082.