

Data Analytics Approach for Short-term Sales Forecasts Using Limited Information in E-commerce Marketplace

Christopher Chin Fung Chee ¹ , Kang Leng Chiew ¹ , Izzatul Nabila bt Sarbini ¹ ,
Eileen Kho Huei Jing ² 

¹ Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Kota Samarahan, Malaysia

² Karuna (Sarawak) Enterprise Sdn. Bhd., Malaysia

Corresponding author: Kang Leng Chiew (klchiew@unimas.my)

Abstract

E-commerce has become very important in our daily lives. Many business transactions are made easier on this platform. Sellers and consumers are the two main parties that gain a lot of benefits from it. Although many sellers are attracted to set up their businesses on this online platform, it also causes challenges such as a highly competitive business environment and unpredictable sales. Thus, we propose a data analytics approach for short-term sales forecasts using limited information in the e-commerce marketplace. Product details are scraped from the e-commerce marketplace using a content scraping tool. Since the information in the e-commerce marketplace is limited and essential, scraped product details are pre-processed and constructed into meaningful data. These data are used in the computation of the forecasting methods. Three types of quantitative forecasting methods are computed and compared. These are simple moving average, dynamic linear regression and exponential smoothing. Three different evaluation metrics, namely mean absolute deviation, mean absolute percentage error and mean squared error, are used for the performance evaluation in order to determine the most suitable forecasting method. In our experiment, we found that the simple moving average has the best forecasting accuracy among other forecasting methods. Therefore, the application of the simple moving average forecasting method is suitable and can be used in the e-commerce marketplace for sales forecasting.

Keywords

Electronic commerce; Data analytics; Sales forecast; Simple moving average; Regression.

Citation: Chee, C. C. F. C., Chiew, K. L., Sarbini, I. N., & Jing, E. K. H. (2022). Data Analytics Approach for Short-term Sales Forecasts Using Limited Information in E-commerce Marketplace. *Acta Informatica Pragensia*, 11(3), Forthcoming articles. <https://doi.org/10.18267/j.aip.196>

Academic Editor: Zdenek Smutny, Prague University of Economics and Business, Czech Republic

Copyright: © 2022 by the author(s). Licensee Prague University of Economics and Business, Czech Republic.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (CC BY 4.0).

1 Introduction

A marketplace is a particular place where buyers and sellers come together to interact directly and indirectly to exchange goods and services for money. A market involves a process which enables demand and supply to operate and determine prices agreed by both buyers and sellers. In the early stage before the Internet explosion, marketplaces were physical commercial locations such as shopping malls, tele-shopping via TV, door-to-door, etc. Buyers and sellers typically met physically to promote products and services and negotiate prices with each other.

In the late 1990s, the Internet changed the method of people doing commerce. People can purchase their preferred products and services through the Internet. The main reason that people tend to shop online is because it saves time and energy compared to traditional shopping (Naseri et al., 2021). Besides, consumers can receive the purchased products directly at their door. This convenient purchasing method increases the usage of e-commerce platforms. High usage of these platforms has increased the volume of trade in the e-commerce marketplace exponentially from year to year (Najmi et al., 2015). Apart from that, it has been reported that 50 percent of Malaysians (16.53 million) are online shoppers (International Trade Administration, n.d.). The statistics shows that the e-commerce platform has created a very significant impact on our daily life.

The increasing number of sellers in the e-commerce marketplace has created a situation where the market is crowded with many privately labelled items of similar products. The volume and variety of products in the e-commerce marketplace greatly increases with the increasing numbers of sellers. In order to compete for sales among the sellers, a seller may cut down revenue and sell the product at a lower price. By using this approach, the seller may attempt to get a lot of sales and gain market share, but the seller cannot earn much profit to support the business in the long term. Some sellers even practice selling at a loss to gain market share. Therefore, it is important for a seller to choose the right product (e.g., product in high demand and relatively less competition). In fact, this situation can be improved if the seller can perform some sales forecasting such as a short-term forecast for a product before selling it. This is definitely helpful for a seller to compete in a crowded marketplace, especially for new sellers.

Besides competition, lack of meaningful data on the platform interface is another issue. Although the digitalisation of the marketplace generates many data on product details such as product name, specifications, manufacturer, distributor, etc., these data are insufficient for sellers to identify profitable products. Sellers cannot use the data directly to make comparisons. One of the main reasons is that the data are considered raw data and meant for buyers instead of sellers. The data shown on the interface are not representative if used independently. Thus, such data are less useful and it is very difficult for sellers to make an informed decision in choosing the right product.

In order to solve these problems, product research tools are introduced. A product research tool is a necessary tool for sellers to venture into the e-commerce marketplace. This tool can be used to filter and shortlist products in high demand, which will help sellers narrow down and find the most potential and suitable product. Some example of this product research tool include Jungle Scout¹, ZonGuru², Viral Launch³ and CamelCamelCamel⁴. Although these tools are very useful, they are platform-specific and proprietary. For example, JungleScout is only applicable to Amazon.com. It would be helpful if there was

¹ See, <https://www.junglescout.com/>

² See, <https://www.zonguru.com/>

³ See, <https://viral-launch.com/>

⁴ See, <https://camelcamelcamel.com/>

an open model which would be applicable to different platforms. To the best of the authors' knowledge, there is no other similar research paper that focuses on the product research method in the literature.

We understand that short-term sales forecasts are necessary for highly competitive marketplaces. However, sufficient historical data are required to perform accurate forecasting. The accuracy of sales forecasting will be affected by the limited information in the digitalised marketplace. Therefore, limited information on each product (especially quantitative data) is studied to make good use of it and to ensure accurate forecasting results. In this study, we propose a data analytics approach for short-term sales forecasting using limited information in the e-commerce marketplace. We are aware that the product research tool is a complete system that can also be used to discover potential products. However, the core of the product research tool is the computational method used to discover potential products. Therefore, our focus is on the computational method, which is the quantitative forecasting approach.

In recent years, many quantitative forecasting models have been developed and each of them has its special use. To carry out a short-term forecast using limited information in the e-commerce marketplace, a forecasting model must be computationally simple and accurate. This is because sellers often require quick and accurate results from the forecast to plan ahead before making any important decisions. Hence, three type of quantitative forecasting methods, namely simple moving average (SMA), linear regression (LR) and exponential smoothing (ES), are investigated and compared in this study. In order to determine a forecasting model with the highest accuracy when using limited information, evaluation metrics such as mean squared error (MSE), mean absolute deviation (MAD) and mean absolute percentage error (MAPE) are applied. The rest of this paper is organised as follows: Section 2 presents the literature review, which focuses on data mining and other research on quantitative forecasting. Section 3 discusses the proposed methodology in detail. Section 4 describes the experimental results. Finally, Section 5 highlights the contribution of our work and concludes the paper.

2 Literature review

Data mining is a process that extracts hidden information from a larger set of raw data. The data mining technique can help us study the trend and patterns of collected data. It gives us a big picture of the data that we might want to study. Thus, decision makers can make better decisions based on analysing the results of data mining instead of acting blindly. It is very common to see data mining applied in the area of marketing. For example, data mining is used to extract useful and hidden information to support marketing decisions (Bach & Alessa, 2014). According to Bach & Alessa (2014), credit card marketers make better decisions and target the right customers with the help of a system that employs the data mining technique. In another example, marketers from a lending organisation had integrated data mining to form a system called credit scoring. The system is used to overcome the complex and unstructured credit risk evaluation (Kambal et al., 2013). By using this tool, marketers can identify the credit risk and minimise non-performing loans to avoid unnecessary losses.

Data mining is also used in forecasting. Forecasting is a process that predicts a future occurrence or event using historical data (Hanggara, 2021). For example, sellers perform forecasting to foresee possible problems such as overstocking, stockout, demand and sales issues that happen frequently in many businesses. Forecasting can improve a company's stocking flexibility. Company owners can change their strategy and allocate resources accordingly. For example, they can reduce the stock of a product when they have forecasted that the product is not in high demand, and invest in other products in higher demand. With such a forecasting approach, the company can stay flexible and react to different possible risks in order to reduce unnecessary losses. This is because high profit and large flexibility of business operation often relates to precise forecasting (Ren et al., 2020). Therefore, forecasting is an essential technique to help a company be more operationally flexible and grow.

One of the challenges in sales forecasting is the rapidly changing and complicated business environment in the e-commerce marketplace (Zhao & Wang, 2017). Such a business environment will affect the accuracy of sales forecasts and cause wrong decisions to be made. In order to deal with it, a short-term sales forecast should be carried out. Although there is limited research on short-term sales forecasts using limited information in the e-commerce marketplace, short-term forecasts of passenger demand under on-demand ride services could explain the importance of short-term forecasts in a rapidly changing environment. An accurate short-term passenger demand forecast is necessary to help taxi drivers and on-demand ride service platforms overcome unhealthy scenarios such as oversupply and overdemand (Ke et al., 2017). According to Ke et al. (2017), short-term forecasts are able to shorten passengers' waiting times while optimising the utilisation rate of the drivers. Such evidence suggests that short-term sales forecasts may be suitable for the rapidly changing e-commerce marketplace in order to optimise inventory management and sellers' profit.

Common forecasting used in businesses frequently relates to a time series quantitative forecasting model. This model can predict the future value based on trends and patterns of past data. A time series quantitative forecasting model is applicable in many fields as long as there exist past data in a time series. The application of different time series quantitative forecasting methods will be discussed.

The application of simple moving average (SMA) to forecast car demand in Indonesia during the outbreak of COVID-19 pandemic helps the automotive industry reduce unnecessary overproduction of vehicles (Hanggara, 2021). According to Hanggara (2021), the automotive industry was affected when the public was instructed to stay at home during the pandemic. Another study claiming that SMA is simple but efficient for determining trends in the marketplace is by Nadhira, Gadisku & Peranginangin (2021). According to Nadhira et al. (2021), sales demand of the personal care product Softex 1400-M can be forecasted using SMA to prevent stockouts or overstocks.

Apart from SMA, linear regression is another type of a time series quantitative forecasting model. A study from Bangladesh used linear regression in crime investigation (Awal et al., 2016). Previous crime data were used to train the linear regression model to forecast the future trend of crime. The authors forecasted the trend of crime that includes dacoity, robbery, burglary and theft.

Exponential smoothing is another type of time series quantitative forecasting model. It smooths the time series data using a smoothing factor to reduce fluctuated forecast values. The smoothing factor is a value between 0 and 1. A higher smoothing factor will result in lesser smoothing and higher responsiveness to variations in the historical data. In Sinaga & Irawati (2020), exponential smoothing was used to forecast medical disposable supply demand. Smoothing factors (α) of 0.1 and 0.3 were used for the computation of exponential smoothing. They used previous actual demand data as the input for the forecasting model. They achieved superior performance by using α of 0.1. Besides, another study by Aini et al. (2017) used exponential smoothing to forecast the sales of 3kg LPG cylinders. The sales data were fed into the forecasting model with α values of 0.1, 0.5 and 0.9. The authors concluded that the α values of 0.9 has the least error in their experiment.

Accuracy evaluation is an important process which is used to measure and understand the quality of a forecasting model. It indicates the performance of a forecasting model in a specific area. There are several common evaluation metrics used to measure the accuracy, which include mean square error (MSE), mean absolute deviation (MAD) and mean average percentage error (MAPE) (Nadhira et al., 2021). MSE and MAD share similar properties as both metrics are the mean of totalled errors. The only difference between MSE and MAD is that MSE is the mean of totalled squared errors while MAD is the mean of totalled absolute errors. A higher value of MSE and MAD indicates that the forecasting result is less accurate compared to a lower value. On the other hand, MAPE shows the prediction percentage error compared to the actual value (Khair et al., 2017). Although using any of the evaluation metrics alone can achieve the purpose of evaluation, using a combination of them can often help cross-check the experiment

results and remove uncertainty. Such a combination can be found in Kilimci et al. (2019), where MAPE and MAD were combined for the second decision integration strategy to determine the accuracy of the forecasting models.

3 Methodology

The proposed methodology involves three main phases: data collection, data formation and quantitative forecasting. First, we use a content scraping tool to collect historical sold quantity data for each product from an e-commerce platform. After that, raw data are pre-processed to yield the historical sales data during the data formation phase. Data formation phase consists of two sub-phases: data pre-processing and data construction. After that, the historical sales data are used in the quantitative forecasting phase. During this phase, different quantitative forecasting methods are involved in the computation to generate forecasted sales. Lastly, the forecasted results are computed. Figure 1 shows the general architecture of the proposed methodology.

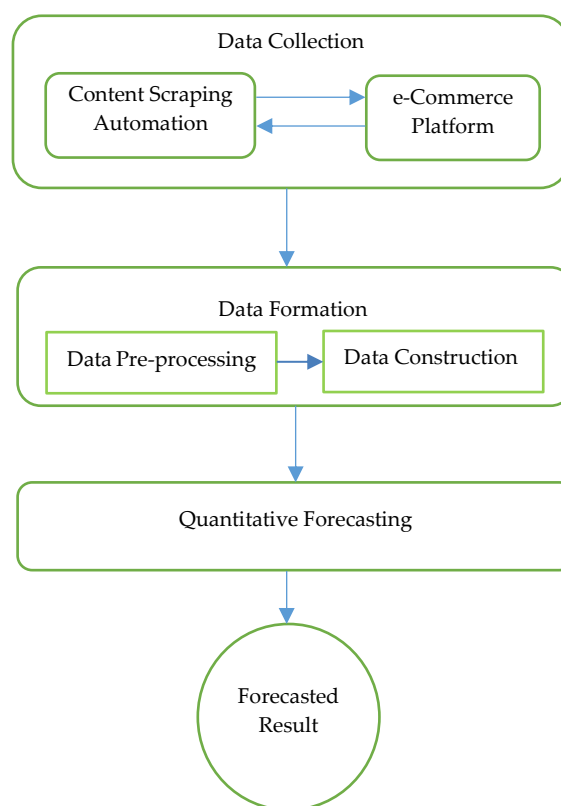


Figure 1. Architecture of proposed system.

3.1 Data collection

In this paper, the Shopee Malaysia⁵ e-commerce platform was selected among other existing e-commerce platforms. This is because the Shopee website provides useful raw data, which is the sold quantity of each product. This sold quantity data record the accumulated sales quantity, which can be transformed to form the historical sales and used for the forecasting. In order to extract these data, an automated scraping module was developed using Selenium Automation⁶. During the scraping process, the module extracted the data from the HTML webpages and stored them in the database.

⁵ See, <https://shopee.com.my/>

⁶ See, <https://www.selenium.dev/>

Shopee has a huge catalogue with many categories. Covering the whole catalogue for this paper might be infeasible due to the time constraint. It was necessary to narrow and focus on a category. Thus, products from the Mobile & Accessories category of Shopee were selected in this study. Under this category, 6 sub-categories of Audio, Cables & Chargers, Cases & Covers, Powerbanks & Batteries, Screen Protectors and Selfie Accessories were chosen. The reason to choose this category is based on the suggestion from Cheriyan & Tamilarasi (2019). Before the scraping process started, 300 products from each category were selected randomly. This random selection created a pool of 1800 products. Since not all 1800 products were relevant to our study, an exclusion process was required. Foreign products (outside Malaysia), live shows and payment services which are not considered tangible products were excluded. After the exclusion process, 100 products from each category were chosen randomly from the remaining valid products. Each of these selected products was used as the seed for the daily scraping processes. Therefore, a total of 600 products were scraped daily during the data collection phase. The data used for this paper are data scraped from 10 January 2022 to 28 February 2022.

3.2 Data formation

The collected raw data were pre-processed and constructed during the data formation phase. There are two operations during pre-processing phase. The first operation is to convert the scraped raw data such as *Sold Quantity* to a number format. For example, '1.3k' is converted into a whole number of 1300. The second operation is to handle invalid data. Essential information such as *Sold Quantity* is important for the forecasting and cannot be left empty. Thus, it is best to pre-process the data with an invalid data replacement procedure. In our proposed replacement procedure, the invalid data were initialised with the value of their previous *Sold Quantity*. The replacement of previous data would overcome the issue of discontinuity of *Sold Quantity*. This workaround is reasonable because it mimics the situation where there are no *Daily Sales* for that particular day. Please note that invalid data may occur during the data collection. For example, the developers of the e-commerce website may periodically update their website, and this may cause some data to no longer exist. Other reasons include slow webpage loading and rendering, which will trigger a timeout of the scraping tool. Table 1 shows the error difference of raw and pre-processed partial sample data.

Table 1. Error difference of raw and pre-processed data.

Raw data			Pre-processed data		
<i>Sold Quantity</i>	<i>Forecasted Sold Quantity</i>	Error	<i>Sold Quantity</i>	<i>Forecasted Sold Quantity</i>	Error
6500	6550	(50)	6500	6550	(50)
0	6700	(6700)	6500	6700	(200)
6900	6900	0	6900	6900	0
7100	7000	100	7100	7000	100
Total absolute error		6850	Total absolute error		350

It is clearly shown that the error due to discontinuity of *Sold Quantity* can be reduced if invalid data are replaced by the previous data. Since *Sold Quantity* is the accumulated daily sales quantities, we can derive the *Daily Sales* by calculating the difference between current and previous *Sold Quantity*. This calculation is conducted during the data construction phase. Note that the data collection during the scraping process is organised daily; we label each scraped data item as *Scrape-k* (i.e., *k* is a running number). The calculated *Daily Sales* were used as the input for the computation of the forecasting model.

3.3 Time series quantitative forecasting

From the previous phases, we know that *Sold Quantity* is the limited information in the e-commerce platform. After extracting *Daily Sales* (time series data) from the *Sold Quantity*, we used different time series quantitative forecasting models to compute the forecasting results and identify the best model. The time series quantitative forecasting models are simple moving average (SMA), linear regression (LR) and exponential smoothing (ES). Besides, the forecasting model should be equipped with short-term forecasting characteristics in order to help sellers compete on the market. Previous data should be excluded and are not taken into account for computation. Therefore, the moving window concept is applied to the existing LR forecasting model. The modified forecasting model that applies the moving window is dynamic linear regression (DLR). The moving window is a technique that takes a subset of data in a fixed amount, adding a new value to the subset and removing an old value from the subset simultaneously while shifting across time. Research shows that better smoothness can be achieved with filtered noise when the window size is increasing (Raudys & Pabarskaite, 2018). So, a moving window of 10 days was used throughout this experiment.

3.3.1 Simple moving average

Simple moving average (SMA) is a value-smoothing forecasting method that computes the average as the output from a group of input values (Pataropura et al., 2019). The formula for SMA is as follows:

$$SMA_i^w = \frac{1}{w} \sum_{k=i}^{i-1+w} A_k, \quad i = 1, 2, 3, \dots, n - w \quad (1)$$

where A_k is the data, n is the total number of the data and w is the window size.

The group of input values to the SMA calculation are those data enclosed within the moving window at the current step. As the moving window shifts forward, the first value within the window is excluded and a new one in the data list is included in the window. This moving window calculation is continued until the end of the list.

3.3.2 Dynamic linear regression

Linear regression (LR) is a forecasting method that predicts the value of dependent variables as a linear function of an independent variable (Awal et al., 2016). The formula of linear regression is as follows:

$$y = \alpha + \beta x \quad (2)$$

where y is the dependent variable, x is the independent variable, n is the number of data used, β is the slope and α is the y-intercept. α and β are obtained using the following equations:

$$\alpha = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \quad (3)$$

$$\beta = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (4)$$

In this paper, the dependent variable y refers to *Daily Sales* while the independent variable x refers to *Scrape number*. LR uses the least squares method to fit a straight line to the data. The value of each point

to form the straight line is computed using the past data. The estimated value can be determined by observing the value along the straight line.

Based on the concept of linear regression, the slope is constant across time. This implies that the rate of change of *Daily Sales* is always the same. However, time series data have fluctuation characteristics. Fluctuation means irregular rises and drops in the data values. Therefore, a dynamic approach is required to best fit the fluctuation characteristics. In order to achieve a dynamic adaptability, a moving window mechanism is incorporated into the LR. This enhanced method is called dynamic linear regression (DLR). DLR has the same formula as LR with the exception that the inputs used for each computation are based on a moving window. When the moving window of DLR is shifted forward at every step, a localised slope of the regression is produced based on the data points within the moving window. These localised slopes generate some fluctuating forecasting results that may match the fluctuating pattern of time series forecasting. This is a desired characteristic as it can serve as a correction mechanism for over- and under-forecasting. Figure 2 shows a comparison of linear regression and dynamic linear regression.

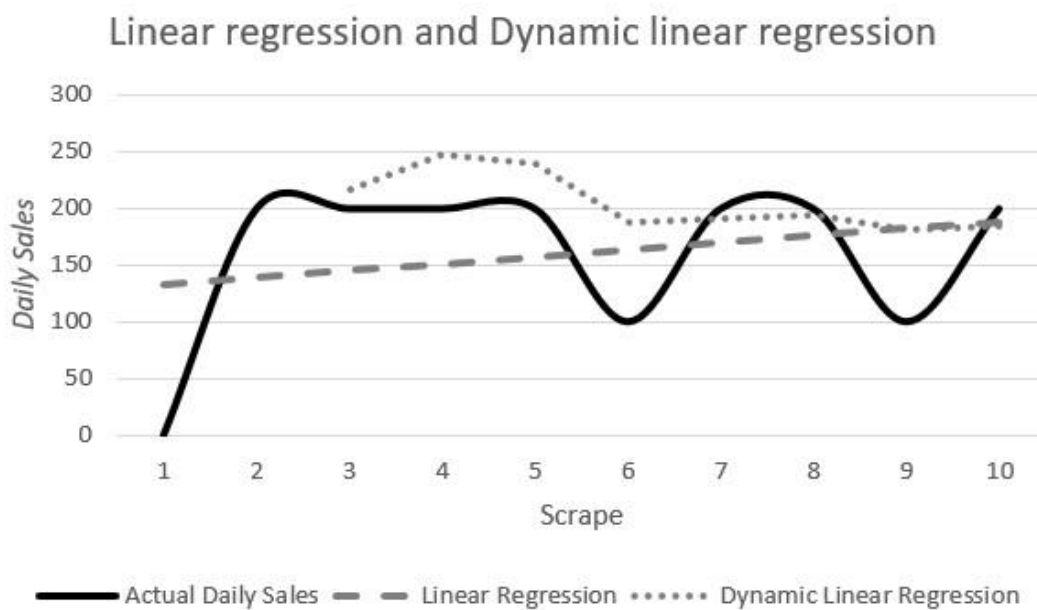


Figure 2. Comparison of linear regression and dynamic linear regression.

Based on Figure 2, the *Daily Sales* of LR are increasing at a steady rate while the *Daily Sales* of DLR show a fluctuating pattern. This is because DLR smooths the values to match the actual data. It is clear that the forecasting pattern of DLR fluctuates based on recent data. This mechanism can correct over- and under-forecasting. Therefore, DLR can adapt to the fluctuating pattern of time series data and forecast based on recent data. Table 2 shows the application of dynamic linear regression using the sample data.

Table 2. Example of forecasting using dynamic linear regression with moving window of 3.

Scrape	Sold Quantity	Daily Sales	Forecasted Daily Sales	Forecasted Sold Quantity
1	6500	0	-	-
2	6700	200	-	-
3	6900	200	216	-
4	7100	200	247	7316
5	7300	200	240	7347

Scrape	Sold Quantity	Daily Sales	Forecasted Daily Sales	Forecasted Sold Quantity
6	7400	100	187	7540
7	7600	200	191	7587
8	7800	200	194	7791
9	7900	100	181	7994
10	8100	200	184	8081
11		-		8284

3.3.3 Exponential smoothing

Exponential smoothing (ES) is a technique that uses a variable decreasing exponentially over time in the exponential function for time series data smoothing (Sinaga & Irawati, 2020). The formula for ES is as follows:

$$ES_i = \alpha A_{i-1} + (1 - \alpha)F_{i-1} \quad (5)$$

where A_{i-1} is the previous *Daily Sales*, F_{i-1} is the previous *Forecasted Daily Sales* and α is the smoothing factor.

The smoothing level of ES is determined by the α value. A smaller α value produces a higher smoothing level of ES. When a smaller α value is used, it causes ES to be less responsive to the variations in historical data. In this paper, ES with α value of 0.1, 0.5 and 0.9 were used. These three values were chosen to determine the effects of different smoothing factors on sales forecasting and were denoted as ES1, ES5 and ES9, respectively. Figure 3 shows the forecasting results using different smoothing factors.

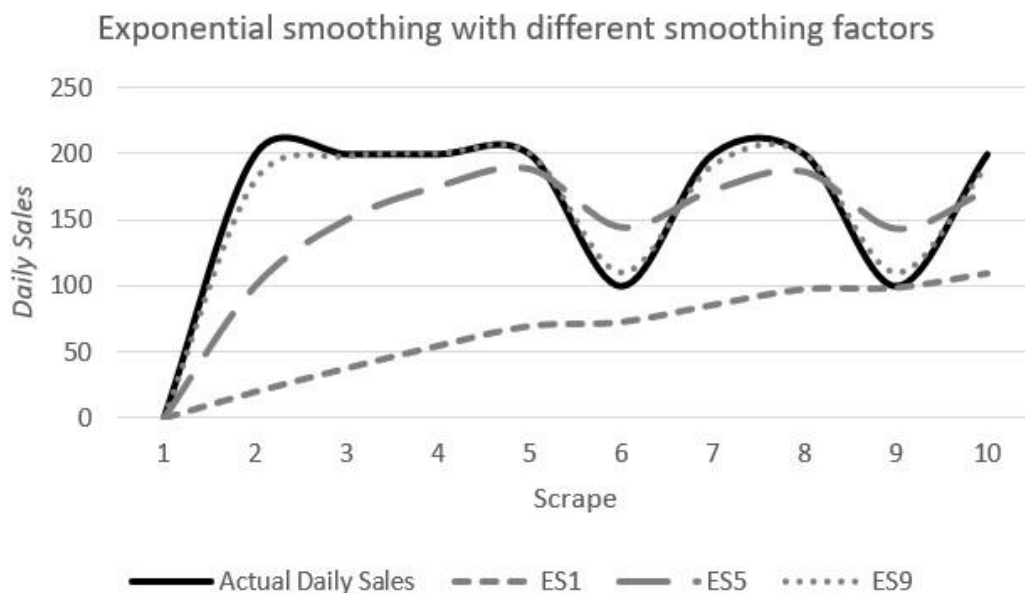


Figure 3. Comparison of different smoothing factors.

Table 3 shows an example of forecasting using exponential smoothing with an α value of 0.9 based on the sample data.

Table 3. Example of forecasting using exponential smoothing with $\alpha = 0.9$.

Scrape	Sold Quantity	Daily Sales	Forecasted Daily Sales	Forecasted Sold Quantity
1	6500	0	0	-
2	6700	200	180	6500
3	6900	200	198	6880
4	7100	200	200	7098
5	7300	200	200	7300
6	7400	100	110	7500
7	7600	200	191	7510
8	7800	200	200	7791
9	7900	100	110	8000
10	8100	200	191	8010
11		-		8291

3.4 Evaluation metrics

In order to evaluate the performance, we used MAD, MAPE and MSE. By using MAD, we can directly know the error difference. For example, if the MAD is 5, it means the forecasted result has deviated ± 5 units from the actual result. The lower the error, the higher the accuracy. As discussed in 3.1 above, a total of 600 different products were selected for the experiment. The performance of each product was evaluated using MAD, MAPE and MSE. The evaluation results were generated for each scraping cycle. In order to obtain an overall result, the MAD, MAPE and MSE of different products in the same scraping cycle were averaged to determine the average of different mean errors. Therefore, the averaged MAD, averaged MAPE and averaged MSE were used as the final evaluation metrics in this study.

4 Experimental results

Figure 4 illustrates the averaged MAD of different forecasting methods from Scrape-15 to Scrape-50. From the figure, we can see that the averaged MAD of all the forecasting methods increase over time. Although there is a sudden spike in Scrape-20, this situation is not abnormal as long as other forecasting models rise and drop together at the same time. This is because *Daily Sales* can suddenly rise and drop due to some special circumstances such as promotion. The sudden spike could be possibly due to the upcoming Chinese New Year 2022 promotion.

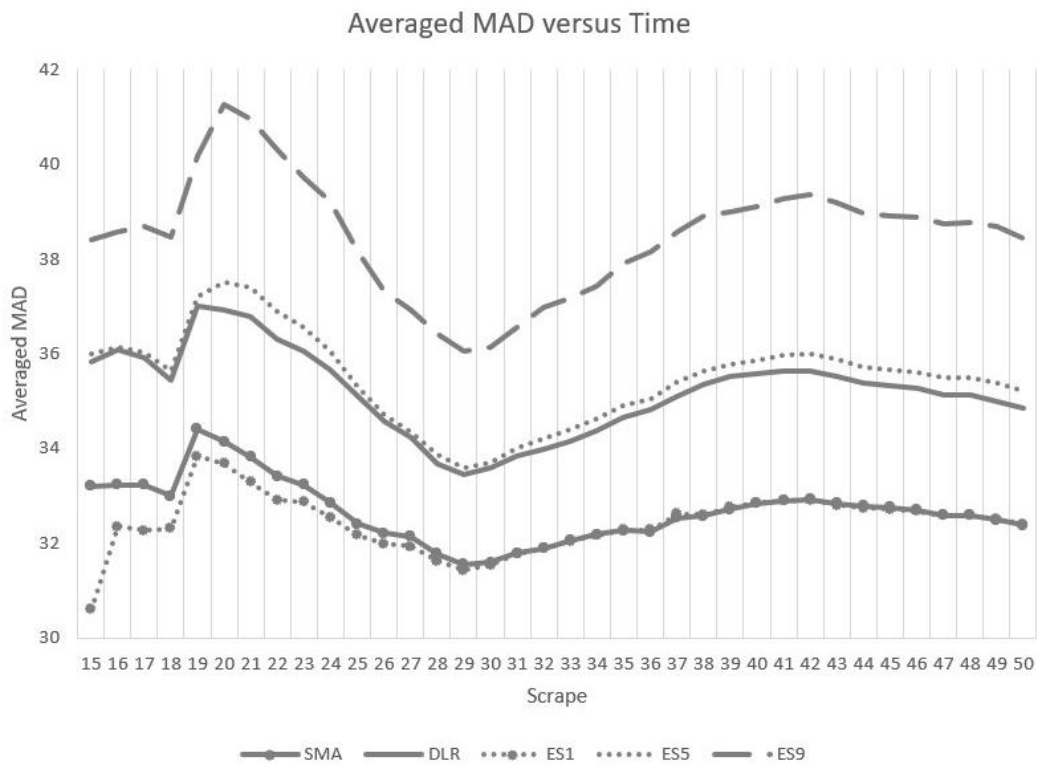


Figure 4. Forecasting performance evaluation using averaged MAD.

Figure 5 illustrates the averaged MAPE of the different forecasting methods from Scrape-15 to Scrape-50. From the figure, we can see that the pattern of the line in Figure 5 is similar to that in Figure 4 for each of the forecasting models. Based on Figure 4, it is clear that ES1 and SMA have similar averaged MAD after Scrape-30. Figure 5 clearly shows that the SMA has the least averaged MAPE after Scrape-30.

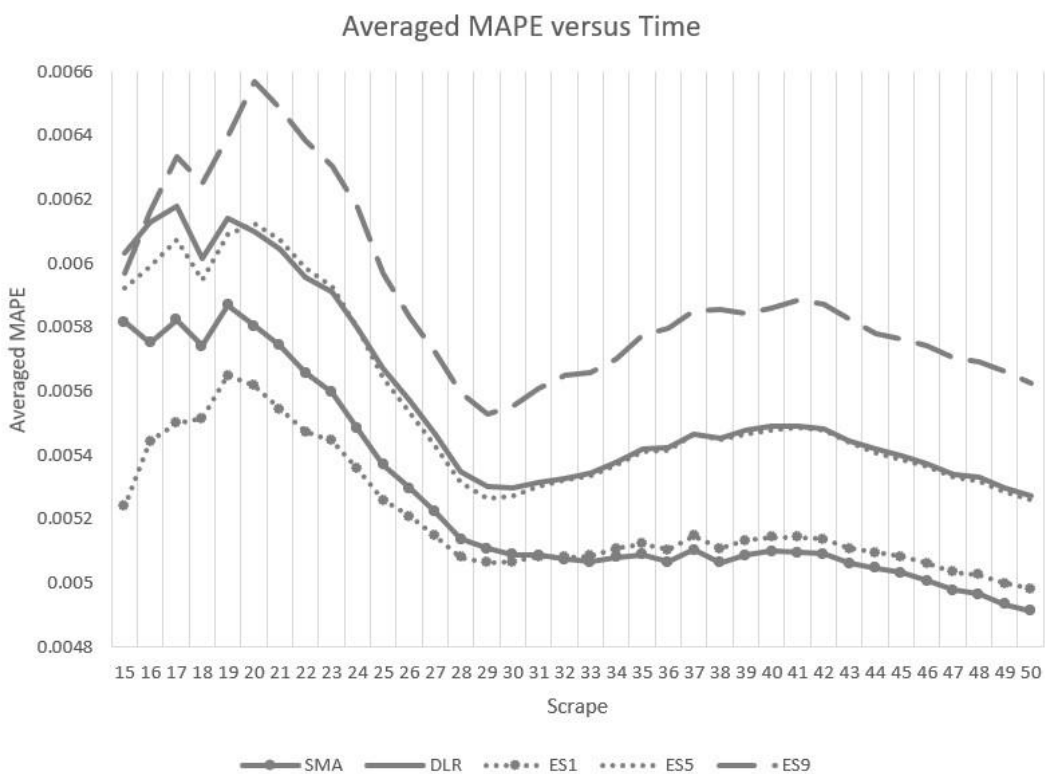


Figure 5. Forecasting performance evaluation using averaged MAPE.

In order to further verify that SMA has the least error, another evaluation metric was used. Figure 6 shows the averaged MSE of the different forecasting methods from Scrape-15 to Scrape-50. Figure 6 shows that SMA outperformed the other forecasting models and recorded the least averaged mean squared errors throughout the experiment.

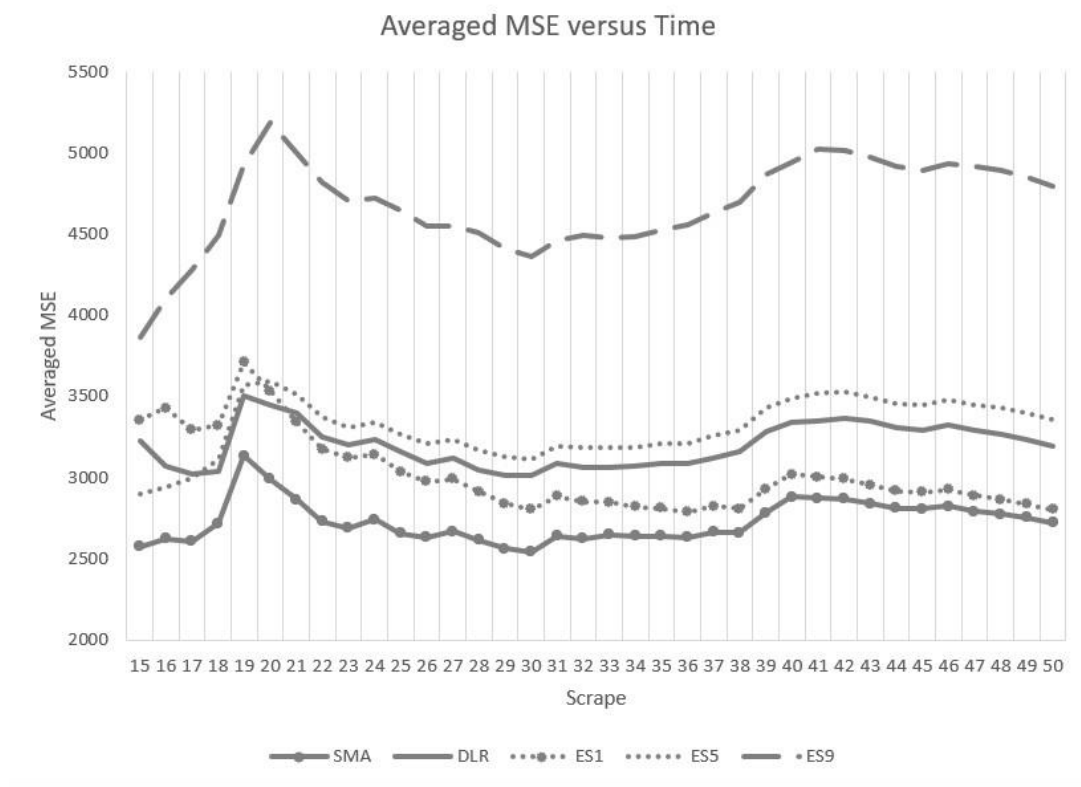


Figure 6. Forecasting performance evaluation using averaged MSE.

Based on the results shown in Figures 4, 5 and 6, the performance of SMA is superior compared to the other forecasting models. Therefore, SMA is the proposed time series quantitative forecasting model for short-term sales forecast using limited information.

To further illustrate the effectiveness of the proposed method, we selected randomly one of the products that was not tested in the above experiments to evaluate the forecasting result. The product is *Macaron Stereo Wired Earphone*. Since the process involves many data, we will list out the comparison from Scrape-15 to Scrape-20 only for the ease of presentation. Table 4 shows the *Forecasted Sold Quantity* and *Actual Sold Quantity* for *Macaron Stereo Wired Earphone*.

Table 4. Forecasted Sold Quantity of different models and actual Sold Quantity for Macaron Stereo Wired Earphone.

Forecasting model	Forecasted Sold Quantity					
	Scrape 15	Scrape 16	Scrape 17	Scrape 18	Scrape 19	Scrape 20
SMA	61,370	61,670	61,860	62,160	62,460	62,880
DLR	61,318	61,635	61,816	62,138	62,460	62,896
ES1	61,439	61,735	61,922	62,220	62,518	62,926
ES5	61,319	61,660	61,830	62,165	62,482	62,941
ES9	61,300	61,690	61,809	62,191	62,499	62,990
Actual	61,400	61,600	61,900	62,200	62,600	62,800

Table 5 shows the *Absolute Error* and *Relative Error* for *Macaron Stereo Wired Earphone*. These errors were computed from the differences between actual and forecasted *Sold Quantity* for each model from Table 4.

Table 5. Absolute and relative error for Macaron Stereo Wired Earphone.

Forecasting model		Error						Sum
		Scrape 15	Scrape 16	Scrape 17	Scrape 18	Scrape 19	Scrape 20	
SMA	Absolute	30	70	40	40	140	80	400
	Relative (%)	0.0005	0.0012	0.0007	0.0007	0.0023	0.0013	0.0067
DLR	Absolute	82	35	84	62	140	96	499
	Relative (%)	0.0014	0.0006	0.0014	0.0010	0.0023	0.0016	0.0083
ES1	Absolute	39	135	22	20	82	126	424
	Relative (%)	0.0007	0.0022	0.0004	0.0004	0.0014	0.0021	0.0072
ES5	Absolute	81	60	70	35	118	141	505
	Relative (%)	0.0014	0.0010	0.0012	0.0006	0.0019	0.0023	0.0084
ES9	Absolute	100	90	91	9	101	190	581
	Relative (%)	0.0017	0.0015	0.0015	0.0002	0.0017	0.0031	0.0097

Based on Table 5, we found that SMA has the least absolute and relative error. This result is also in line with the evaluation results using averaged MAD, averaged MAPE and averaged MSE in Figures 4, 5 and 6, respectively. The absolute error and relative error of SMA from Scrape-15 to Scrape-20 are 400 and 0.0067%, respectively.

5 Conclusion

In this paper, scraped quantitative data were pre-processed and constructed into appropriate data that could be used for a forecasting experiment. In order to identify a suitable forecasting model for the e-commerce marketplace, we conducted several experiments using different forecasting models: simple moving average, dynamic linear regression and exponential smoothing. As a result, simple moving average is the most suitable forecasting model because it showed the least averaged mean absolute deviation, averaged mean absolute percentage error and averaged mean squared error when we compared using different models.

The simple moving average forecasting model uses a direct averaging calculation that consumes less computational resources and can return the forecasting result quickly. Hence, it is more practical for use as a forecasting model as sellers often require quick information in the competitive e-commerce marketplace.

Additionally, the forecasting model using simple moving average is one of the main components of more comprehensive software for sellers, namely product research tools. In order to fully utilise the proposed forecasting model in the e-commerce marketplace, the proposed forecasting model could be developed as a plugin module for product research tools. Sellers can use the forecasting feature of the product research tool to find potential products to be added to their selling catalogue.

Without a forecasting module, a product research tool can only show static information such as product rating or specification. Clearly, this is incomplete as the level of demand for the product is missing. Selling a low-demand product would be difficult for sellers to compete in a crowded e-commerce marketplace. On the other hand, sellers can identify the performance of a product on the market by using the forecasting

feature of the product research tool. Sellers can forecast the *Daily Sales* of a product, potentially determine and compare the demand for different products. Therefore, it is essential to be able to forecast using the simple moving average forecasting model as the quantitative approach before making a decision on selecting the right product to sell.

As future work, larger dataset will be included to explore a big data approach. We also plan to include more data on product details such as reviews, ratings and prices to improve the forecasting model. This could help increase the feature dimension that produces a more accurate forecasting model. This improved forecasting model will be suitable for extension to a comprehensive product research tool. Besides more data, more variety in product categories within an e-commerce platform can also be considered. Increasing the data space will provide more learning attributes and will be beneficial to computational approaches such as machine learning and big data.

Additional Information and Declarations

Funding: The funding for this project is made possible through the research grant obtained from UNIMAS under the Smart Partnership Grant Scheme [Grant No: F08/PARTNERS/2121/2021]

Conflict of Interests: The authors declare no conflict of interest.


Author Contributions: C.C.F.C.: Methodology, Investigation, Data collection, Writing – Original draft. K.L.C.: Methodology, Validation, Supervision, Writing – Review & Editing. I.N.S.: Supervision, Writing – Review & Editing. E.K.H.J.: Supervision.

Data Availability: The data that support the findings of this study are available from the corresponding author.

References

- Aini, A. N., Purbaningtyas, R., & Adityo, R. D. (2017). Forecasting Sales 3kg LPG using Single Exponential Smoothing Method. *Journal of Electrical Engineering and Computer Sciences*, 2(2), 281–286.
- Awal, M. A., Rabbi, J., Hossain, S. I., & Hashem, M. M. (2016). Using Linear Regression to Forecast Future Trends in Crime of Bangladesh. In *2016 5th International Conference on Informatics, Electronics and Vision* (pp. 333-338). IEEE. <https://doi.org/10.1109/ICIEV.2016.7760021>
- Bach, C., & Alessa, A. (2014). Data Mining and Knowledge Management for Marketing. *International Journal of Innovation and Scientific Research*, 2(2), 321–328.
- Cheriyian, A., & Tamilarasi, D. S. (2019). A Study on the Influence of Mobile Accessories Purchased Impulsively, with Special Reference to Chennai City. *International Journal of Advanced Science and Technology*, 28(17), 780–790.
- Hanggara, F. D. (2021). Forecasting Car Demand in Indonesia with Moving Average Method. *Journal of Engineering Science and Technology Management*, 1(1), 1–6.
- International Trade Administration. (n.d.). Malaysia – eCommerce. <https://www.privacyshield.gov/article?id=Malaysia-E-Commerce>
- Kambal, E., Osman, I., Taha, M., Mohammed, N., & Mohammed, S. (2013). Credit Scoring Using Data Mining Techniques with Particular Reference to Sudanese Banks. In *2013 International Conference on Computing, Electrical and Electronic Engineering (ICCEEE)* (pp. 378-383). IEEE. <https://doi.org/10.1109/ICCEEE.2013.6633966>
- Ke, J., Zheng, H., Yang, H., & Chen, X. (Michael). (2017). Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging Technologies*, 85, 591–608. <https://doi.org/10.1016/j.trc.2017.10.016>
- Khair, U., Fahmi, H., Hakim, S. A., & Rahim, R. (2017). Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error. *Journal of Physics: Conference Series*, 930, Article no. 012002. <https://doi.org/10.1088/1742-6596/930/1/012002>
- Kilimci, Z. H., Akyuz, A. O., Uysal, M., Akyokus, S., Uysal, M. O., Atak Bulbul, B., & Ekmis, M. A. (2019). An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain. *Complexity*, 2019, 1–15. <https://doi.org/10.1155/2019/9067367>

- Nadhira, A. T., Gadisku, C. A., & Peranginangin, S. M.** (2021). Demand Forecasting Comparison of Softex 1400-M using Single Moving Average Method and Single Exponential Smoothing Method. In *Proceedings of the International Conference on Industrial Engineering and Operations Management*, (pp. 452–459). IEOM Society International.
- Najmi, E., Hashmi, K., Malik, Z., Rezgui, A., & Khan, H. U.** (2015). CAPRA: a comprehensive approach to product ranking using customer reviews. *Computing*, 97(8), 843–867. <https://doi.org/10.1007/s00607-015-0439-8>
- Naseri, R. N., Hussin, H., Esa, M. M., Aziz, N. E., & Nordin, M. N.** (2021). What is a Population in Online Shopping Research? A perspective from Malaysia. *Turkish Journal of Computer and Mathematics Education*, 12(4), 654–658. <https://doi.org/10.17762/turcomat.v12i4.549>
- Pataropura, A., Riki, R., & Saputra, A.** (2019). Sales Analysis Using the Forecasting Method. *Bit-Tech*, 1(3), 146–149. <https://doi.org/10.32877/bt.v1i3.79>
- Raudys, A., & Pabarskaite, Z.** (2018). Optimising the Smoothness and Accuracy of Moving Average for Stock Price Data. *Technological and Economic Development of Economy*, 24(3), 984–1003. <https://doi.org/10.3846/20294913.2016.1216906>
- Ren, S., Chan, H.-L., & Siqin, T.** (2020). Demand forecasting in retail operations for fashionable products: methods, practices, and real case study. *Annals of Operations Research*, 291, 761–777. <https://doi.org/10.1007/s10479-019-03148-8>
- Sinaga, H. D., & Irawati, N.** (2020). A Medical Disposable Supply Demand Forecasting By Moving Average and Exponential Smoothing Method. In *Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA)*. EAI. <https://doi.org/10.4108/eai.24-1-2018.2292378>
- Zhao, K., & Wang, C.** (2017). Sales Forecast in e-Commerce using Convolutional Neural Network. arXiv:1708.07946v1. <https://doi.org/10.48550/arXiv.1708.07946>

Editorial record: The article has been peer-reviewed. First submission received on 26 July 2022. Revision received on 4 October 2022. Accepted for publication on 18 October 2022. The editor in charge of coordinating the peer-review of this manuscript and approving it for publication was Zdenek Smutny .

Acta Informatica Pragensia is published by Prague University of Economics and Business, Czech Republic.

ISSN: 1805-4951
