

PriceCop – Price Monitor and Prediction Using Linear Regression and LSVM-ABC Methods for E-commerce Platform

Mohamed Zaim Shahrel

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
E-mail: zaimshahrels@gmail.com

Sofianita Mutalib and Shuzlina Abdul-Rahman

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
E-mail: sofi@fskm.uitm.edu.my, shuzlina@fskm.uitm.edu.my

Received: 24 September 2020; Revised: 23 October 2020; Accepted: 26 November 2020; Published: 08 February 2021

Abstract: In early 2020, the world was shocked by the outbreak of COVID-19. World Health Organization (WHO) urged people to stay indoors to avoid the risk of infection. Thus, more people started to shop online, significantly increasing the number of e-commerce users. After some time, users noticed that a few irresponsible online retailers misled customers by hiking product prices before and during the sale, then applying huge discounts. Unfortunately, the “discounted” prices were found to be similar or only slightly lower than standard pricing. This problem occurs because users were unable to monitor product pricing due to time restrictions. This study proposes a Web application named PriceCop to help customers’ monitor product pricing. PriceCop is a significant application because it offers price prediction features to help users analyse product pricing within the next day; thus, it can help users to plan before making purchases. The price prediction model is developed by using Linear Regression (LR) technique. LR is commonly used to determine outcomes and used as predictors. Least Squares Support Vector Machine (LSSVM) and Artificial Bee Colony (ABC) are used as a comparison to evaluate the accuracy of the LR technique. LSSVM-ABC was initially proposed for stock market price predictions. The results show the accuracy of pricing prediction using LSSVM-ABC is 84%, while it is 62% when LR is employed. ABC is integrated into SVM to optimize the solution and is responsible for the best solution in every iteration. Even though LSSVM-ABC predicts product pricing more accurately than LR, this technique is best trained using at least a year’s worth of product prices, and the data is limited for this purpose. In the future, the dataset can be collected daily and trained for accuracy.

Index Terms: Price Prediction, Linear Regression, Least Squares Support Vector Machine (LSSVM), Artificial Bee Colony (ABC).

1. Introduction

The story of online shopping began four decades ago, and the platform has steadily grown along with technologies and innovations. The pioneer of e-commerce was Amazon, a company that started by selling books in the early 1990s [1]. According to [2], online retailers nowadays have taken advantage of the advancement of information technology as retailers can track and analyse customers’ private information, such as their behaviour and preferences. In theory, analysing customer behaviour and preferences enable online retailers to discriminate prices by offering discounts, vouchers, and advertisements [2]. Price discrimination within the customer base is still widely practiced. Although a policy exists in online retail where “one price for all” should be applied, online retailers may not adhere to it [1,2]. The practice of e-commerce price discrimination was further revealed when one of Amazon’s customers disclosed how the price of a DVD sold on the platform changed after he deleted the cookies on his personal computer. Amazon publicly apologized and refunded all customers who had paid more for the product [2].

Market manipulation is a type of abuse where an attempt is made to interfere with an existing free and fair trade by creating a false or misleading price of a product in the market. Market manipulation is prohibited in most countries. Currently, a consumer may prefer to shop online because it is more reliable and convenient to browse any product with a variety of prices. Similar to a conventional shop, an online store also offers discounts and vouchers to attract customers. However, behind all these offers, an online retailer can trick buyers through sudden price hikes and hidden costs that add to the end price of the product. This is as demonstrated in the previously quoted case of the Amazon

customer who realized the product prices displayed on the platform were misleading. The sudden price hike is an example of a price manipulation scheme imposed by online retailers leading to customers paying more than they should. Moreover, most consumers prefer to use online retail prices. According to Hinz et al. [2], consumers most likely do not use the Manufacturer's Suggested Retail Price (MSRP) as their preference, resulting in them not acknowledging the base price for a product. Thus, the risk of consumers being misled is high when they use online retail price as a product's base price.

Price manipulation happens when consumers cannot monitor prices at all times; thus, online retailers take advantage of this weakness to change prices without notice. Several attributes have been observed to tackle this problem, such as time, geolocation, personal preference, and MSRP. According to Zuiderveen & Poort [3], different geolocation, time, and a consumer's personal preferences affect the prices shown on their e-commerce website. Reference price represents the base price of a product. Most product prices are set by the MSRP [4], and MSRP will help consumers to compare the product prices against the retail prices set by retailers. In the context of e-commerce, online retailers are responsible for setting the prices of their products based on the MSRP. Most prices are set either lower or on par with the MSRP. According to Hwang & Kim [5], online retailers are allowed to set their product prices above the MSRP, but they should price their products below base prices to attract customers. Online retailers should estimate their cost and revenue before setting the price of a product.

The main objective of this research is to study the following prediction techniques, which are Linear Regression (LR) and Least Squares Support Vector Machine – Artificial Bee Colony optimization (LSSVM-ABC), to predict product pricing in the market using previous prices. The current solution to predict product pricing is based on the market conditions covering supply and demand. It will help the customers to be notified with the price of products in interest. The customers can check with the offered price in the ecommerce website, whether it is a reasonable price or it might be price hike. At the end of this research, a Web application is proposed by integrating one of the techniques to predict prices on the Amazon platform.

2. Related Works

In a marketing study, an early discussion mentioned the fundamentals of pricing a product in the market. Thus, several past methods and opinions will be discussed in this section. Previous research suggests that machine learning approaches are commonly required to prepare and train the dataset to discover any abnormality in market pricing on e-commerce.

2.1 Analysis on Product Pricing Method

According to Soon [6], in the past year, pricing a product is recognized as a mandatory tool used to maximize profit. It can be applied either in revenue management or supply chain management, or in daily operations industries to manipulate demand or regulate the production of goods and services. Extensive research has been done in the past decades to determine many pricing strategies, including dynamic and static strategies, single and multiple product strategies, and competitive strategies. There are three types of pricing models: static non-competitive pricing models, dynamic non-competitive pricing models, and competitive pricing models. The first, static non-competitive pricing models, suggest that a product must have a fixed price even when sold by others. This method has a higher tendency to monopolize the market with a single price. The second, dynamic non-competitive pricing models, refer to the strategy where the pricing of a product may change based on trends or demands over time. The price can increase if there is a higher demand but an insufficient supply of goods and decrease if there is less demand but an excess of stock. The second situation is usually referred to as inflation. The third, competitive pricing model, suggests that the seller has the right to price their product depending on their objectives and decisions. The seller can also reduce their price as a strategy to compete with others. Commonly, to apply this model, a lot of attributes or variables need to be observed to avoid loss and improve profit.

2.2 Analysis on Price Discount Method

Lin et al. [7] stated that a successful marketing strategy will produce a successful business model. The most effective and direct approach to this strategy is by offering consumers to ask for a discount. The most popular business model, Name-Your-Own-Price (NYOP), is a massive success in facilitating sales. Usually, when a seller plans to reduce their prices, they will review successful trades in the past and the timeline. The foundation in discounting a product is to reduce the price after the target sale has been achieved within a specific period. The second strategy in reducing prices to attract customers is a clearance sale. To clear shelves of unsold stock, sellers will often offer their products at a discount. The reason the product was not bought in the first place might have been because the customer thought the price was unworthy of the product, or they thought there was a better-value alternative elsewhere. Thus, sellers will reduce the price either close to or past cost value. Typically, the sellers will have already accounted for the loss before significantly discounting the product [8].

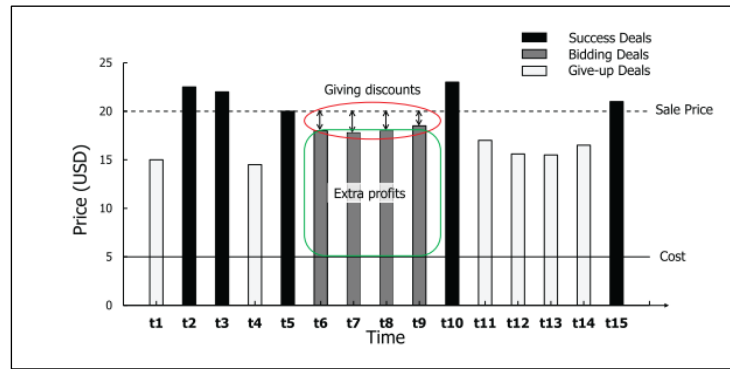


Fig.1. Observation on seller giving discount on the product [7].

Moreover, there is also a price discount policy concerning inventory cost to protect buyers from seller manipulation [9]. There are disagreements with how suppliers continue to assume buyers’ optimal behaviour to formulate terms for quantity discount-pricing schedule [9]. The formulas are derived for price that will maximise economic gain for the supplier from the price revision, buyer gains, and sum of gains for both supplier and buyer. It is suggested that supplier may have taken advantage of the buyer through pre-determined price, mutual benefits, and incremental discount in the system.

2.3 Analysing Dynamic Pricing Algorithm on Market

Both conventional and online retail sellers need to set prices for their products, and thus, they require a pricing algorithm for their prices to compete in the market. Hwang & Kim also stated that retailers use either an automated algorithm to price their products or an analysed data based on the current or previous market [4].

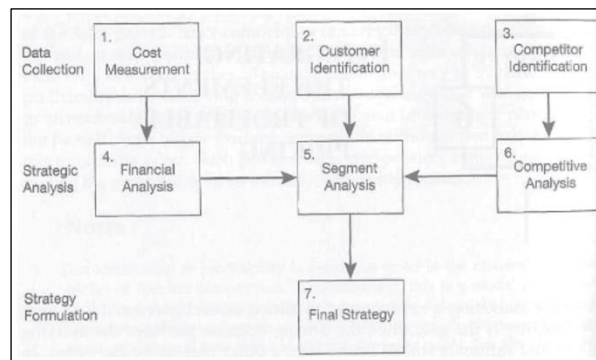


Fig. 2. Steps for more profitable pricing [4].

Hwang & Kim add that there are steps to set pricing to be more profitable [4]. Fig. 2 exhibits the three phases retailers need to go through: data collection, strategic analysis and strategy formulation. In the first phase, data collection, cost is measured through a simple calculation using traditional accounting method, customers’ desires are ascertained, and competitors, either individuals or companies, are identified. In the second phase, strategic analysis, the first step is to analyse the financial status: cost of goods, and revenue that are used to calculate the profit. Then, customers are grouped into different classes based on their locations, socioeconomic status, purchase characteristics and other categories in a segment analysis. Next, the competition is analysed where sellers identify their market competitions based on the products they sell. In the final phase, strategy formulation, sellers need to summarise and analyse all the input gathered from the previous phases and finalize their strategy to maximise profit.

2.4 Other relevant factors

A few past researches mention that retailers purposely conceal hidden costs on their product’s end-price. According to Belleflamme & Vergote [10], in the United States, taxes are not applicable on an online shopping platform, but a few retailers incorporate the taxes onto their product’s end-price, with no mention about it in the final price. This unethical behaviour affects customers as they end up spending more money without realizing it. Leibbrandt stated that online market can be affected by this unethical practice as customers’ trust would erode over time [11]. This was proven in a previous case in China, according to Gao & Wu, where most buyers in China have reservations making purchases on an online shopping platform if the price differs from any other source of pricing [12].

Monopolization still occurs on e-commerce platforms even though there are multiple sources. This happens because larger retailers on the online market have price control of the product. Belleflamme & Vergote found that, this

is common when major retailers control the product price if they also distribute the product to other retailers [10]. These major retailers are able to sell the product cheaper because they do not need to include any additional cost in their product price. However, the other retailers have to include their cost in their product price. Thus, these larger retailers can monopolize the market as they can control the product price.

Profiteering is a term for the act of making excessive or unfair profit illegally. In the context of e-commerce, profiteering still occurs. According to Romero & Willburn, in the early stages of e-commerce growth, a few retailers profited by incorporating non-existing taxes and costs onto their product's final price [13]. Based on the research, consumers pay for the product without being aware of how the product is priced as the bill does not include the full breakdown. A user's online browsing behaviour is commonly monitored either by a data analyst or the website's server host. Although most websites state that consumer or user privacy is protected, the risk of data leakage still exists [3]. Belleflamme & Vergote state that online retailers set the price differently based on location, cookies, and time [10]. Hinz et al. add that consumers are aware that different discount offers, and vouchers are given among users. This proves that consumer behaviour is monitored based on their preferences [2].

3. Supervised Predictive Methods

Supervised Machine Learning employs algorithms for instances to produce general hypotheses, then makes predictions on future instances. Supervised classification is one of the most used by intelligent systems. A few of the machine learning algorithms are Random Forest (RF), Naïve Bayes (NB), Support Vector Machine (SVM), Neural Networks (Perceptron), and Decision Tree (J48). The classification problem is the standard formulation of the supervised learning task as the learner is required to learn the function for mapping a vector into one of several classes by overview at previous examples. A set of rules from instances can be learned by inductive machine learning in order to create a classifier to be used for generalizing a new instance. [14] [15] The process of applying supervised machine learning is described in Fig. 3 below.

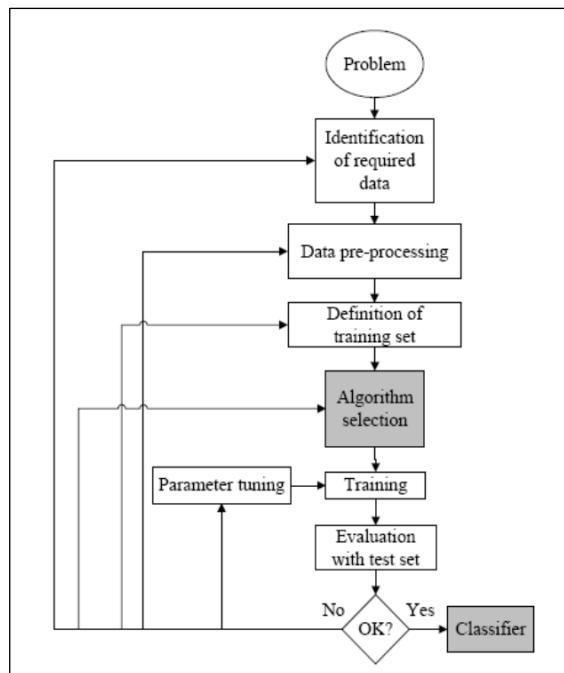


Fig.3. Process of Supervised Machine Learning [14].

To analyse complex data and data rich in semantics, as well as design an intelligent information system, artificial intelligence (AI) technique is required to process the data collected. Supervised machine learning is considered an AI technique. A variety of area applications is due to the broad range of opportunities when using an AI technique. Nowadays, the neural network method is the most broadly known for analysing data [15], [16]. A brief explanation of the methods used in the project, which are linear regression (LR) and Least Squares Support Vector Machine (LSSVM), is given in the next subsections.

3.1 Linear Regression

Linear regression is an approach under supervised learning. Supervised learning is a learning function that generalizes function from the dataset and dependent learning. Linear regression is a technique to test the relationship

between two variables, where the relationship is expressed in this technique in the form of an equation. Linear regression also helps in analysing the strength between two dependent variables and also predictor variables [17]. Furthermore, linear regression can also be used to predict a pattern or the price of a product. To determine an accurate estimation of price or pattern, a number of regularization methods in linear regression need to be applied, such as ridge, lasso, and their relatives [18].

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (1)$$

Equation (1) presents an equation used for linear regression [17]. The equation has the form of:

$$Y = a + bX \quad (2)$$

where,

Y , is a dependent variable that represents the Y-axis,

X , is an independent variable that represents the X-axis,

b , is the slope of the line and a is the y-intercept.

As demonstrated in (1), ε represents a random error as there is always an error term or residual term.

3.2 Least Squares Support Vector Machine

Least squares support vector machines (LSSVM) are least squares version of support vector machine (SVM). SVM is used for classification and regression analysis, as well as a set of supervised learning methods that analyse data and for pattern recognition, while LSSVM is used to find solutions by solving a set of linear equations instead of convex quadratic programming (QP) problems. This method is proposed by Suykens and Vandewalle [19].

The goal of LSSVM is to construct function $f(x) = y$, as output y_i represents the dependency on input x_i . This function is formulated as:

$$f(x) = W^T \varphi(x) + b \quad (3)$$

Where, W and $(x) : R^p \rightarrow R^n$ are $n \times 1$ column vectors, and $b \in R$.

The main differences between LSSVM and SVM are that LSSVM includes the equality of constraints instead of inequalities, and it is based on the least squares cost function. LSSVM algorithm solves similar minimization problems that are found in the SVM method as displayed in Fig. 5. Moreover, the LSSVM method also solves linear problems while SVM solve quadratic problems.

Throughout the method, by using,

$$y_i = W^T p(x_1) + b + e_i \quad (4)$$

as equality constraints of LSSVM, a matrix is obtained as displayed in Fig. 4 below.

$$\begin{bmatrix} I & 0 & 0 & -Z^T \\ 0 & 0 & 0 & -1^T \\ 0 & 0 & CI & -I \\ Z & 1 & I & 0 \end{bmatrix} \begin{bmatrix} W \\ b \\ e \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ y \end{bmatrix}$$

Fig.4. Matrix obtained using LSSVM method [19].

Then, by defining,

$$k = zz^T \quad (5)$$

and,

$$\lambda = c^{-1} \quad (6)$$

parameter as kernel matrix, the following overall solution is optimized by the condition.

$$\begin{bmatrix} 0 & 1^T \\ 1 & k + \lambda \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (7)$$

3.3 Optimization for Predictive Analytics

According to Hertog, predictive analytics appear to be used more often in practice compared to prescriptive analytics [20]. However, the use of optimization in predictive analytics is arguable. He adds that most of the predictive analytics application does not exploit the use of optimization as it can be a huge value-add for the practice. In this topic, the fundamentals of Particle Swarm Optimization and Artificial Bee Colony Algorithm are discussed as they are used in this project to compare the accuracy of Linear Regression technique. Particle swarm optimization (PSO) was first proposed by Eberhart and Kennedy in 1995. This method is inspired by a simulation of birds flocking to achieve global optimization. Each particle in the PSO method adjusts itself accordingly to get the best position among all the particles in the swarm. After the particles in the swarm converge to the global best point, the iteration will be terminated. When the particle is searching its space, it has its position (X) and velocity (V) [21]. PSO is the easiest implementation and a fast convergence compared to other evolutionary algorithms, such as artificial neural networks, genetic programming, genetic algorithms, and ant colony optimization. It is also applicable to solve complex and complicated problems.

3.4 Artificial Bee Colony Algorithm

The Artificial Bee Colony (ABC) algorithm was introduced by Karaboga in 2005 for parameter optimization. ABC was inspired by the intelligent behaviour of honey bees and recognized as a recent swarm intelligence technique [22], as shown in Fig. 5. There are three groups of honeybees: employed, onlooker, and scout bees. These groups of bees form a colony of artificial bees. Half of the colony is composed of employed bees, while the other half is composed of onlooker and scout bees. The fundamental of this algorithm is that each employed bee is responsible for one nectar source. The objective of the whole colony is to maximize the amount of nectar collected. Solutions are represented by the duty of employed bees searching for food sources, and fitness values are represented by the amount of nectar. Onlooker bees waiting in the hive receive the information obtained by employed bees, and these onlooker bees will then decide the nectar source to be exploited based on the information obtained. This group will also determine the allocation of employed bees as scout bees. This new group will then find new valuable food sources. They randomly search for sources near the hive [19].

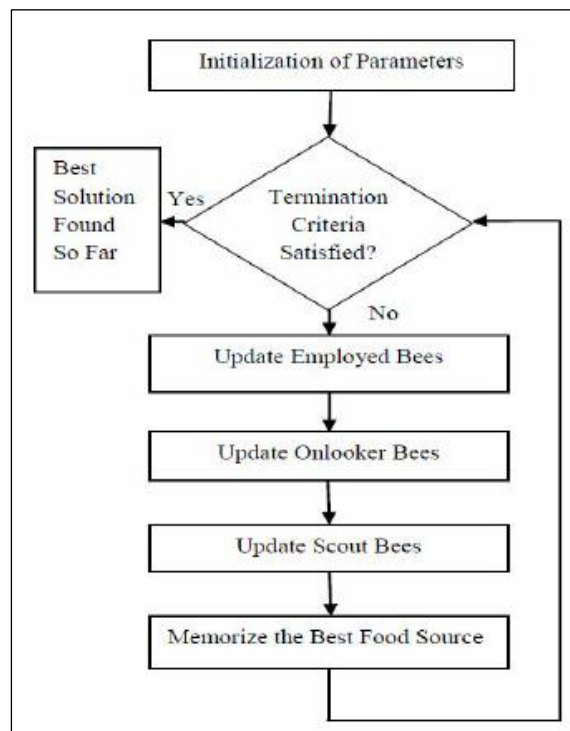


Fig.5. Artificial Bee Colony algorithm flow [22].

4. Research Process

To predict the product pricing using Linear Regression (LR) and Least Squares Support Vector Machine – Artificial Bee Colony (LSSVM-ABC), several processes, data collection and pre-processing, model training and testing, and Web deployment are discussed in the subsections.

4.1 Data Collection and Pre-processing

During this phase, the collection of data is carried out by using the web scraping method. The data collected on this website is the product name, price, and seller. In this study, the highest-grossing product is selected as it enables market prediction to be made more readily. The idea is to collect the same product from different sellers.

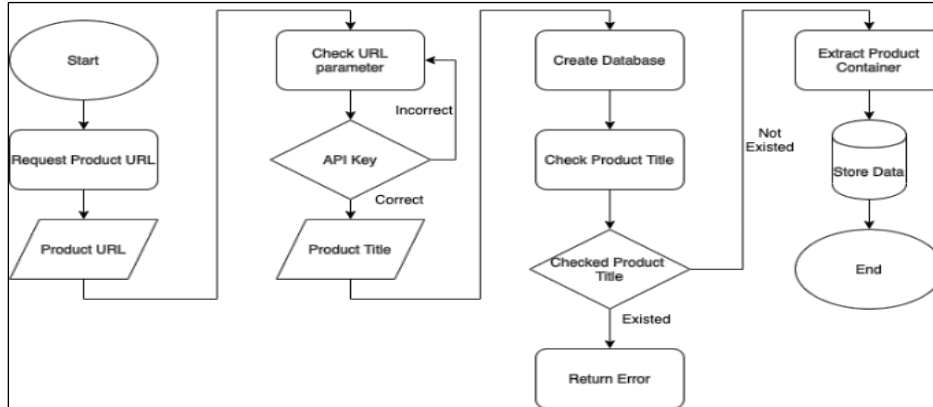


Fig.6. Data collection flow using web scraping.

Fig. 6 depicts a flow chart of data scraping from an e-commerce website. The language used for this method is Python. At the early stage of data scraping, the programme requires the product URL, which is entered by the user as input. Then, the URL parameter is checked to confirm its correctness. This stage is vital as the programme can only accept a URL from an e-commerce website. Any other URLs that do not come from an e-commerce website will be ignored or declined. Next, an API key for data scraping from the website has to be verified. The process will continue once the API key is confirmed. Otherwise, the API key has to be replaced with the correct one.

These steps are required as the programme needs to create a connection with the scraping service provider. ScrapeStack services are used in this study. After the key is verified, the product title is scraped and extracted. The title is then renamed in a class container. Next, the programme will create a MongoDB database, and the database will contain a cluster and collection for storing data. The product title will be checked to see if it already pre-exists in the database. If the product already exists in the database, then the programme will not store the data and will return an error. Otherwise, it will be stored in the database, and the product container (price, picture, and rating) will be extracted. The process will terminate after the data is stored in the database. Data collected in the previous step will be pre-processed. Since the data is collected in real-time upon request, a soft code is implemented in the backend programme. Any character separation and selection will be done while the data is collected. Then, the processed data will be stored in the MongoDB Atlas database server as part of the collection.

The collected attributes are the name of the item, availability, rating from the customer, URL, image, and price. The price of the product is stored in an array that is subject to comparison when the input user initiates a new request for a price check. Scraping data from an e-commerce website is a non-continuous task. Thus, the crawler or bot scraper can avoid getting blocked from the website since a few websites block crawlers if it continuously scrapes data. The data from 10 products were collected throughout this study, as shown in Table 1, beginning from 1st June 2019 and ending on 6th July 2020 when each product has between 365 and 372 of data pricing/ Meanwhile, Table 2 provides price information for a sample of product Corsair M65 Pro RGB Mouse.

Table 1. List of 10 Products' Name.

No.	Product's Name
1.	Apple iPhone X, GSM Unlocked 5.8in, 64 GB - Space Gray (Renewed)
2.	New Apple MacBook Air (13-inch, 8GB RAM, 128GB Storage) - Gold
3.	LG 38WK95C-W 38-Inch Class 21:9 Curved UltraWide WQHD+ Monitor
4.	Borderlands 3 Super Deluxe Edition - PlayStation 4
5.	Acer Predator Helios 300 Gaming Laptop PC, 15.6inch Full HD 144Hz
6.	Apple MMEF2AM/A AirPods Wireless Bluetooth Headset for iPhones
7.	Corsair Gaming M65 PRO RGB FPS Gaming Mouse, Backlit RGB LED
8.	LG 34UM69G-B 34" 21:9 UltraWide IPS Monitor with 1ms
9.	Echo Plus (2nd gen) – Premium sound with built-in smart home hub
10.	Corsair CH-9206015-NA Gaming K55 RGB Keyboard, Backlit LED

4.2 Model Training and Testing

The implementation in the backend is performed using soft code inside the program itself. By using the ‘Sklearn’ Python library, the Linear Regression is successfully implemented and stores the predicted value in the MongoDB database. The attribute selected for this technique is the product price. The data is transferred into an Excel worksheet for visualization and verification. For experimental analysis, the data is also trained and tested with Least Squares Support Vector Machine and Artificial Bee Colony (LSSVM and ABC) technique. This method is implemented using MATLAB software code. The experimental analysis is to compare the accuracy between this technique and Linear Regression.

Table 2. Corsair M65 Pro RGB Mouse pricing and rating for November 2019

Date	Price (CAD)	Rating	Date	Price (CAD)	Rating
01/11/2019	89.99	4.2	16/11/2019	50.46	4.2
02/11/2019	89.99	4.2	17/11/2019	50.46	4.2
03/11/2019	59.8	4.2	18/11/2019	50.46	4.2
04/11/2019	59.8	4.2	19/11/2019	50.46	4.2
05/11/2019	59.8	4.2	20/11/2019	49.9	4.2
06/11/2019	59.8	4.2	21/11/2019	49.9	4.2
07/11/2019	59.8	4.2	22/11/2019	49.9	4.2
08/11/2019	129	4.2	23/11/2019	49.9	4.2
09/11/2019	129	4.2	24/11/2019	49.9	4.2
10/11/2019	129	4.2	25/11/2019	49.9	4.2
11/11/2019	110.89	4.2	26/11/2019	129.9	4.2
12/11/2019	110.89	4.2	27/11/2019	125.89	4.2
13/11/2019	89.99	4.2	28/11/2019	72.99	4.2
14/11/2019	89.99	4.2	29/11/2019	77	4.2
15/11/2019	89.99	4.2	30/11/2019	77	4.2

For experimental analysis, a set of product pricing data for a year is selected for Corsair Gaming M65 Pro RGB Gaming Mouse, as demonstrated in Table 2. Approximately 80% of the data undergoes the training and testing process, while the other 20% is put forward for validation purpose. Fig. 7 portrays the process flow in LSSVM and ABC.

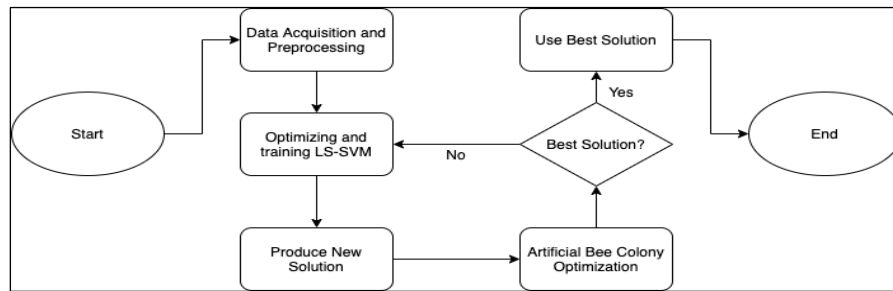


Fig. 7. LSSVM and ABC flow chart.

Fig. 7 portrays the first steps in training, that of data acquisition and pre-processing. The dataset is trained and tested using LSSVM to produce a solution or results of used technique. The solution will then undergo ABC optimization to achieve the best possible price prediction result. At this stage, the best solution with the highest accuracy of prediction is chosen. If the solution is not having the highest accuracy, then it will optimize and train again using LSSVM. These processes are repeated every month in the dataset. Thus, if a new solution is found at the end of the month, then the solution will be used throughout the year for prediction.

To implement LSSVM in the data, a specific formula is used from [19]. The data is assumed as X in the algorithm. Then, X is equal to or assigned to matrix $n \times p$ input data and Y as $n \times 1$ output vector. The trained data set is assigned to:

$$\{x_i, y_i\}_{i=1}^n \quad (8)$$

where,

$$x_1 \in \mathbb{R}^p \quad (9)$$

and,

$$y_i \in R \quad (10)$$

is the goal of LSSVM to construct the function of,

$$f(x) = y \quad (11)$$

Representing the dependence of the output y_i on the input x_i . Then, a kernel needs to be applied based on the data type. Since the data collected is linear, then a suitable kernel is Linear kernel. The kernel is as follows:

$$k(x, x_i) = x_i^T x \quad (12)$$

4.3 Web deployment

The model training and testing are performed in the backend using Python and MATLAB. After the records are processed and trained, the records are transferred into a dashboard. The records will be visualised in a Web application. There are two sections of the web application structure: backend and frontend, as illustrated in Fig. 8. The backend is the server-side of the application where all training/testing processes are done using supervised machine learning, while the frontend is where the communication between the application and the user happens.

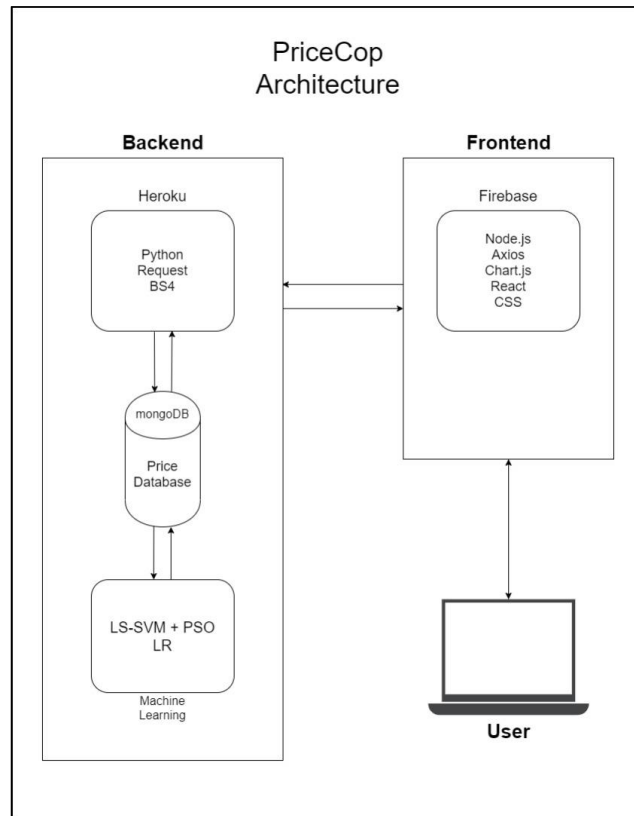


Fig. 8. PriceCop Architecture for Web Deployment.

Fig. 8 illustrates the Python methods and programs hosted on Heroku. The processes performed in the backend are URL request, scraping method, storing and calling data, and machine learning application. The process of scraping method has been discussed in the previous section. The database used for this web application is MongoDB. Any storing and calling data method are done between Heroku and MongoDB server site. MongoDB is a cloud database service provider, and all of the data collection being performed in the backend is automatically stored in a specific array. Heroku is a backend cloud server provider. Since the web application requires continuous communication with the backend, it cannot be hosted on a local server but on the cloud instead. Any request from the frontend is passed on to the backend to process. The user can communicate with the application in the frontend where it contains a user interface and user experience design. Node.js is the main environment that handles and executes any JavaScript code. Axios is used for calling method from the backend where it manages the communication between the backend and frontend.

Chart.js is used to visualize all the graphs on the web application. React native is used to build and handle the framework of the application. CSS is used as colour picker library.

5. Results and Findings

This section focuses on the results and findings of the study based on the selected techniques.

5.1 Linear Regression

The price of Corsair Gaming M65 PRO RGB FPS Gaming Mouse in October 2019 is selected for the result analysis. The graph in Fig. 9 demonstrates the pricing scatter that began to increase on 11th October 2019 and decreased to the original pricing on 20th October 2019. According to the Corsair website, the original price for this product is \$89.99 (CAD). This shows that the seller increased the price of the product above the retail price suggested by the company.

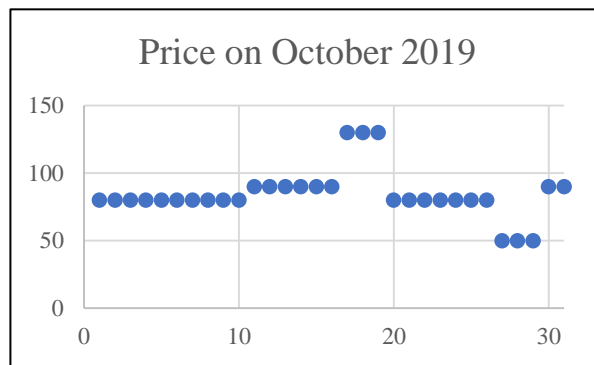


Fig.9. Scatter plot graph for product price on October 2019.

As exhibited in Fig. 10, the x-axis is the product price range, while the y-axis is the date (in November 2019). The price prediction is performed using the Linear Regression technique and this shows that the product price is expected to reduce by the end of the following month, in November 2019. To support this analysis, the Black Friday sale takes place on 27th November every year. Thus, there will be a big sale for every product pricing.

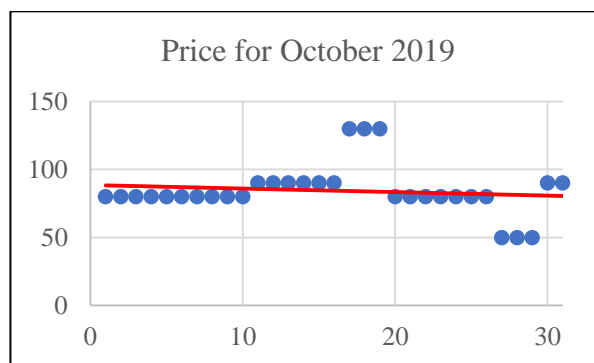


Fig.10. Generated trendline for product price in November 2019 using Linear Regression.

Since soft-coded Python programmes do not show any visualization of prediction data, the data is transferred into an Excel worksheet and analysed using the 'Analysis Toolpak' add-on with comparable results in the Python programme. The result of the Linear Regression implementation is displayed in Table 3 below.

Table 3. Regression statistics result for product Corsair M65 Pro RGB Mouse.

Regression Statistics	
Multiple R	0.12268
R Square	0.01505
Adjusted R Square	0.01255
Standard Error	22.2664
Observation	396

Multiple R from the table above is the same as the Correlation Coefficient that measures the strength of a linear relationship between two variables. The correlation is strong when the value is 1, and vice versa. When the value is 0, there is no relationship between the two variables. R Square in the table is the Coefficient of Determination, which is used to indicate the goodness of fit. It shows how many points fall on the regression line. The R² value is calculated from the total sum of squares or the sum of the squared deviations of means from the original data. Generally, only 95% of the R² value is considered good. From this result, the fitness value is not considered good even though R² only represents the percentage of dependent variables.

The Adjusted R Square from the table indicates the number of independent variables in the model. Since there are only two attributes used, the value of Adjusted R is meaningless as the value is only useful for the multiple regression method. Standard error from the table indicates the precision of the regression analysis. The smaller the number, the better. From the results table, there is only a 22% error, which indicates a small error in prediction. Finally, Observation in the table indicates the number of records that were observed by the model.

Another set of output analysed the variance of the model: Analysis of Variance (ANOVA). Table 4 shows the results of the analysis, where df indicates the number of degrees of freedom associated with the source of the variance. SS is the sum of squares where the smaller the value of Residual SS compared to Total SS, the better the model fits the data; thus, the result in the table shows that the model is suitable for the data. MS in the table indicates the mean square. F in the table demonstrates F statistic or F-test, used for testing the overall significance of the model. Finally, Significance F is the P-value of F, and this value indicates the reliability of the results. Generally, if the Significance F is less than 0.05, the model is suitable for the data, and vice versa. The verdict is that this model fits the data as the value is 0.01457.

Table 4. Analysis on Variance (ANOVA)

ANOVA					
	df	SS	MS	F	Significance F
Regression	1	2984.73325	2984.73325	6.02014	0.01457
Residual	394	195341.59644	495.79085		
Total	395	198326.32969			

5.2 Least Squares Support Vector Machine and Artificial Bee Colony

Another method used in this study is LSSVM and ABC for experimental analysis by comparing the accuracy against Linear Regression. The LSSVM is applied in MATLAB code, the linear kernel is selected, and the result of LSSVM implementation is revealed in Fig. 11 below.

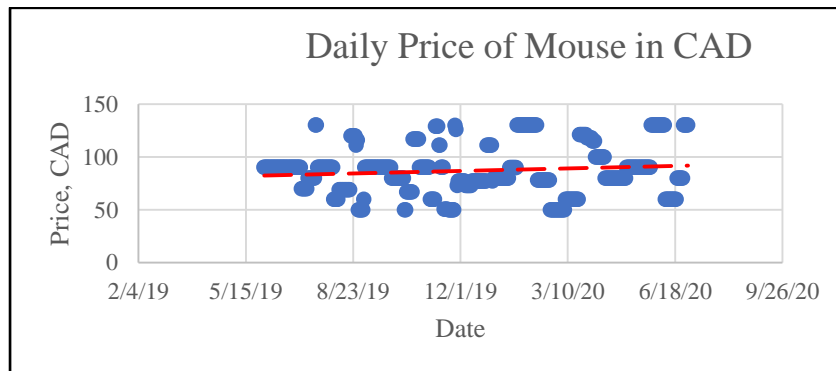


Fig.11. Trendline of LSSVM method for Corsair M65 Pro RGB Mouse price.

As portrayed in Fig. 11, the price is predicted to increase over time until the end of the year. From the data collected, it is comparable to the result since the pricing of Corsair M65 Pro RGB Mouse increases over time. The price only decreases when there is a sale on that date. After implementing LSSVM, Artificial Bee Colony (ABC) optimization is implemented based on the solution. According to Hegasy et al. [19], ABC optimization is responsible for optimizing the prediction model based on the best prediction solution. In implementing ABC optimization, the MATLAB code is initialized by entering the object function value to determine fitness. The formula is as follows:

$$Fit_i = \frac{1}{(1+0.1 \cdot f_{uni})} \quad (13)$$

Then, the trial vector is initialized. Later, the code calculates the probability of the best solution used by employing the following equation:

$$probability = 0.9 * (fit/max(fit)) + 0.1 \tag{14}$$

The goal of this method is to find the best solution by comparing the previously combined solution. Findings show that this method is suitable to predict product pricing as it can adapt or choose the best solution based on the previous combined solution. The price change is dependent on various attributes, such as market trends, sales, the stock market, and the state of the world.

5.3 Accuracy Comparison

Both techniques are applied to compare and test the accuracy of the price prediction. As presented in Fig. 12, the accuracy of applying the LSSVM and ABC method is more accurate than applying Linear Regression because using LSSVM and ABC method provides optimization where only the best result is used throughout the prediction process.

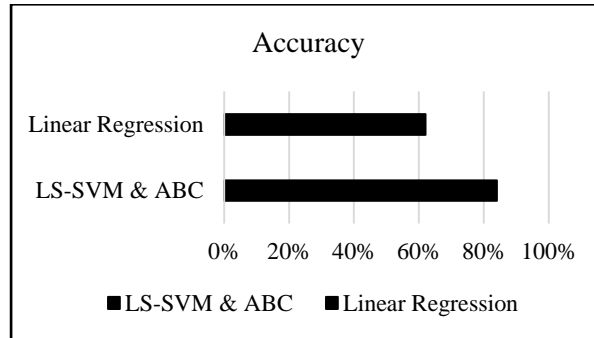


Fig.12. Prediction accuracy comparison between LSSVM & ABC, and Linear Regression

To conclude, it has been found that using LSSVM and ABC is more accurate than using the simple Linear Regression method. However, this method cannot be implemented into the Web application due to some technical constraints even though the accuracy difference between LR and LSSVM-ABC is in the range of 20%.

5.4 Web Deployment

A Web application is created using a JAVA-based application as the frontend, a sample displayed in Fig. 14. The frontend is where the communication between the application and the user happens. The application is hosted locally using Ngrok tunnels. Figure 13 below illustrates the web application design and functionalities. A list of products is added to the MongoDB Atlas database server along with the user email for recording purposes. By adding the email, the user can manage (add or delete) their item list in the web application. Other than that, the user email is also required for the application to notify them of any price change. The email notifier is soft-coded within the programme in the backend. This can be achieved by using the ‘smtplib’ Python library to send the email notifier to the user. The main purpose of this function is to help with user alerts when price changes. Thus, issues related to price manipulation can be solved using this method.



Fig.13. Dashboard from Web browser.

6. Conclusion

This paper presents a Web application namely PriceCop that helps the user or customers monitor price of product listed at e-commerce, so that they are aware of the price changes. Price changes can be both good and bad. It is good when the price is reduced; the item becomes cheaper and more affordable. It is bad when the price is listed with a huge discount, but in reality, the price after the discount remains similar to the normal price. Thus, the most suitable method to achieve this objective is by creating a system where a user can monitor their item of choice. The system needs to be user-friendly and easy to manage. Product pricing prediction is still not prevalent and is less important to potential customers as they may already have their preferred method to predict when the product will be priced low. For example, many buyers already expect that products will be priced lower and cheaper when a sale is happening. However, some online retailers are already prepared to trick and mislead customers while the sale is ongoing. A Web application is proposed in this study to offer a price prediction feature to help users analyse product pricing within the next day by using Linear Regression (LR) technique. To add value to the paper, the study also evaluates the accuracy based on the LR technique and Least Squares Support Vector Machine – Artificial Bee Colony (LSSVM-ABC). LSSVM-ABC was initially proposed for stock market price prediction. The results show that the accuracy of predicting product prices using LSSVM-ABC is 84%, while it is 62% when LR is employed. This is because ABC is integrated into SVM to optimize the solution found. ABC is responsible for finding the best solution in every iteration. Even though LSSVM-ABC predicts product pricing more accurately than LR, this technique is best when trained using at least a year's worth of product prices, and the data is limited for this purpose. In the future, the dataset can be collected daily and trained for accuracy. A graphical price prediction analysis is also recommended to be implemented into the application as it will help the user to understand the pricing trends for the product. Thus, the user can plan when to purchase their desired item.

References

- [1] Miva, "The History Of Ecommerce: How Did It All Begin?" (October 26, 2011). Available at Miva: <https://www.miva.com/blog/the-history-of-ecommerce-how-did-it-all-begin/>
- [2] O. Hinz, I. Hann, and M. Spann, "Price Discrimination in E-Commerce? An Examination of Dynamic Pricing in Name-Your-Own Price Markets." *MIS Quarterly*, vol. 35, no. 1, 2011, pp. 81–98, JSTOR, 2011.
- [3] F. Zuiderveen Borgesius and J. Poort, "Online Price Discrimination and EU Data Privacy Law," *Journal of Consumer Policy*, vol. 40(3), pp. 347–366, 2017.
- [4] T. Thörner, and T. D. Andersson, "Reference Prices in an E-commerce Context" *An experiment with changing consumer prices Reference prices in an E-commerce context*.
- [5] S.B. Hwang and S. Kim, "Dynamic pricing algorithm for E-Commerce," *Advances in Systems, Computing Sciences and Software Engineering - Proceedings of SCSS 2005*, pp. 149–155, 2006.
- [6] W. Soon, "A review of multi-product pricing models." *Appl. Math. Comput.*, vol. 217, pp. 8149–8165, 2011.
- [7] Y.C. Lin, C. H. Huang, C. C. Hsieh, Y. C. Shu and K. T. Chuang, "Monetary discount strategies for real-time promotion campaign," *26th International World Wide Web Conference, WWW 2017*, pp. 1123–1132, 2017.
- [8] M. Armstrong, and Y. Chen, Discount Pricing. *Mpra*, 35423, Paper No. 40041, posted 17. <https://doi.org/10.5897/JAERD12.088>
- [9] N. H. Shah and V. M Dixit, "Price Discount Strategies: A Review," *Revista Investigacion Operacional*, vol. 26(1), pp. 19–32, 2005.
- [10] P. Belleflamme, "Monopoly Price Discrimination and Privacy: The Hidden Cost of Hiding," (October 15, 2015). Available at SSRN: <https://ssrn.com/abstract=2699085> or <http://dx.doi.org/10.2139/ssrn.2699085>
- [11] A. Leibbrandt, "Behavioral constraints on price discrimination: Experimental evidence on pricing and customer antagonism," *European Economic Review*, vol. 121, 2020.
- [12] Y. Gao and X. Wu, "A Cognitive Model Of Trust In E-Commerce: Evidence From A Field Study In China", *Journal of Applied Business Research*, vol. 26, no. 1, pp. 37–44 Jan. 2010.
- [13] P. Romero and J. Willburn, E-Commerce & Taxation. *Graziadio Business Review (Information Management/Technology (IT))*, vol. 2(3), 1999.
- [14] F.Y. Osisanwo, J.E.T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi, J. Akinjobi, "Supervised Machine Learning Algorithms: Classification and Comparison". *International Journal of Computer Trends and Technology (IJCTT)*, vol. 48(3), pp. 128–138, June 2017.
- [15] A.A. Baharudin, S. Mutalib, N.H. Abd Hamid and N.H.A. Hamid, Price Changes Analysis Using Association Rule Mining on Online Shopping Portals," *International Journal of Engineering & Technology*, vol. 7 (4.31), pp. 266–271, 2018.
- [16] L. Ogiela, *Cognitive Information Systems in Management Chapter 3 - Intelligent Computer Data Analysis Techniques*, Academic Press, February 2017, pp. 1–2.
- [17] K. Kumari and S. Yadav, "Linear Regression Analysis Study," *Journal of the Practice of Cardiovascular Sciences*, vol. 4, pp. 36–38, 2018.
- [18] H. Hirose, Y. Soejima, and K. Hirose, "NNRMLR: A combined method of nearest neighbor regression and multiple linear regression." *Proceedings of the 2012 IIAI International Conference on Advanced Applied Informatics, IIAIAI 2012*, 351–356. <https://doi.org/10.1109/IIAI-AAI.2012.76>
- [19] O. Hegazy, S. O. Soliman, and M. Abdul Salam. "LSSVM-ABC Algorithm for Stock Price prediction." *International Journal of Computer Trends and Technology*, 7(2), 81–92. <https://doi.org/10.14445/22312803/ijctt-v7p121>

- [20] D. den Hertog and K. Postek, "Bridging the gap between predictive and prescriptive analytics - new optimization methodology needed," (December, 21 2016). Available at Optimization Online http://www.optimization-online.org/DB_HTML/2016/12/5779.html
- [21] L. Wang, Y. Cheng and D. C. Liu, "New evolution algorithm based on the standard particle swarm optimization," *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*, Taipei, 2011, pp. 110-114, 2011.
- [22] S. Kumar, V.K. Sharma and R. Kumari, "Randomized Memetic Artificial Bee Colony Algorithm," *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, vol. 3(1), pp. 52-62, April 2014.

Authors' Profiles



Mohamed Zaim Shahrel was born on 26th February 1996. Origin from Ipoh, Perak, Malaysia. Undergraduate student from Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia in Bachelor of Information Technology (Hons.) Intelligent System Engineering Faculty of Computer and Mathematical Sciences.



Sofianita Mutalib received her bachelor's degree in 1998 and master degree in 1999 from Universiti Kebangsaan Malaysia, PhD in Information Technology and Quantitative Sciences – Machine Intelligence from Universiti Teknologi MARA. She is a senior lecturer in Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia and also a member of Research Initiative Group Intelligent Systems (RIGIS). She is actively doing research in applied data mining and data analytics, for various area and different types of data. Her interest also has been shown through yearly publications in proceedings and also journals.



Shuzlina Abdul Rahman received her bachelor's degree in Computer Science from Universiti Sains Malaysia in 1996, M.Sc in Information Technology from Universiti Utara Malaysia in 2000 and PhD in Science and System Management – Optimization in Data Mining from Universiti Kebangsaan Malaysia in 2012. She is currently working as Associate Professor at Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA, Malaysia. She is the head of Research Initiative Group Intelligent Systems (RIGIS). Her research interest includes computational intelligence, machine learning and data analytics & optimization.

How to cite this paper: Mohamed Zaim Shahrel, Sofianita Mutalib, Shuzlina Abdul-Rahman, " PriceCop–Price Monitor and Prediction Using Linear Regression and LSVM-ABC Methods for E-commerce Platform", *International Journal of Information Engineering and Electronic Business(IJIEEB)*, Vol.13, No.1, pp. 1-14, 2021. DOI: 10.5815/ijieeb.2021.01.01