



International Journal of Marketing Management

ISSN 2454 - 5007



www.ijmm.net

Email ID: editor@ijmm.net , ijmm.editor9@gmail.com

The Use of Machine Learning for Sneaker Price Forecasting

¹Mr.Dilshad, ² Mohd Suhayb Mohiuddin, ³ Maaz ul haq, ⁴ Mohammed Riyan Hussain Hyderi,

¹Assistant Professor, Department of CSE-AIML, Lords Institute of Engineering & Technology.

²³⁴ Student Department of CSE-AIML, Lords Institute of Engineering & Technology.

Abstract-

As a self-organizing culture, sneaker culture emerged among teens without external influence. With the advent of the Internet and other digital technologies, sports shoes have acquired monetary and aesthetic value via manual speculation, setting off a flood of such speculation in which every individual is involved. The explosive process behind speculative footwear warrants contemplation as a collective activity involving humans. Fry shoes follow a pattern of practical logic and a series of logical steps. To uncover the ever-changing process of convergence from micro to macro behavior, threshold theory is an essential tool. Hence, the sneaker industry may benefit from machine learning-based pricing predictions. This research use stockx as its data set. The research examines the cost of athletic shoes by analyzing data using a machine learning approach. Predictions were made using four different linear regression models. The accuracy of the OLS regression model is determined to be high. Accordingly, the OLS model works well with this data set.

Predicting the price of shoes using machine learning and linear regression

INTRODUCTION

The research indicated that the global sports shoes market's size and growth rate both continued to expand between 2011 and 2019. The shoe market was worth 146.5 billion USD in 2018, up 72.3 billion from 2011. The size of the sports shoe industry continues to grow in 2019, reaching 168.47 billion USD. The sneaker business is projected to reach \$278 billion in 2025, according to one source. To solve AI challenges, machine learning is a useful tool. Machine learning is one approach to AI. Algorithms of the machine learning variety that can automatically examine datasets for regular patterns and then use those patterns to predict data and characteristics that are not yet known. Machine learning has exceptional ties to ratiocinate statistics because of the large amount of statistical theories used in learning algorithms. Machine learning aims to provide easy-to-handle approximation methods. Knowledge Discovery in Databases (KDD), computer vision (CV), natural language processing (NLP), biometric identification, medical diagnosis, search engines, information security, market analysis, face recognition, and autonomous robotics are just a few of the many applications of machine learning. Few studies have examined the use of machine learning for price prediction in the shoe market, despite the fact that this industry deals with money. Consequently, the issue that has to be addressed in this research is how to do an accurate analysis of the stockx dataset and, based on that analysis, what sort of model is best suited for the dataset. Predicting future shoe pricing helps shoe platforms keep costs low and helps shoe retailers study the school market to get more commodities with good value.

TECHNOLOGYINTRODUCTION

Step one: linear regression

Linear regression is a technique in statistical analysis that finds the interdependent reasonable connection between two or more variables by using regression analysis in mathematical statistics [2]. A normal distribution with an average value of zero is used in the formula $y = w'x + e$, where e is the error [3].

A linear regression model takes an n -unit data set and assumes a linear connection between y and the regressors x . This kind of connection is established using an unobserved error variable ϵ . "Noise" is introduced to the linear connection by means of the random variable. Hence, the model takes on this shape:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i = x_i^T \beta + \epsilon_i, \quad i = 1, \dots, n \quad (1)$$

The transpose is denoted by T in this model. The inner product of β and the vectors x_i is represented by $x_i^T \beta$ in the formula.

In matrix notation, these equations are often expressed as

$$y = X\beta + \varepsilon \quad (2)$$

Where

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}, X = \begin{pmatrix} x_1^T \\ x_2^T \\ \dots \\ x_n^T \end{pmatrix} = \begin{pmatrix} 1 & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ 1 & \dots & x_{np} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_p \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix} \quad (3)$$

Using the input as an independent variable, the ordinary least squares approach adds and multiplies additional variables beta to get the final product. The first beta, which is simply termed "beta_1," is used for determining the slope of the function. [5] If we enter 0 as the value, it will essentially tell us the output. The designated name for the second beta is "beta_2". This number stands for the disparity between the increments of the variables, or the coefficients, from [6] to [7].

The ordinary-least-squares (OLS) technique based on error value is the most often used estimate method in regression analysis. To begin estimating using this approach, get the sum of squares of residuals, which is equal to the square of the number of stars and includes all error values:

$$RSS = \sum_{i=1}^n e_i^2 \quad (4)$$

β_0 and β_1 can be calculated using the following formula:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \end{aligned} \quad (5)$$

among $\bar{x}$ is for x average value, while $\bar{y}$ is for y average of.

The following equation may be used to estimate the variance, assuming that the population error value has a constant variance:

$$\hat{\sigma}_\varepsilon^2 = \frac{RSS}{n-2} \quad (6)$$

When you divide the sample size by the number of parameters that need to be estimated, you get the mean square error as a denominator. To the right of these parameter estimations is shown the standard error:

$$\hat{\sigma}_{\beta_1} = \hat{\sigma}_\varepsilon \sqrt{\frac{1}{\sum (x_i - \bar{x})^2}} \quad (7)$$

$$\hat{\sigma}_{\beta_0} = \hat{\sigma}_{\beta_1} \sqrt{\frac{\sum x_i^2}{n}} \quad (8)$$

Researchers will have and using the aforementioned model. The value of is predicted using this formula.

METHODOLOGY

Released by stockxin, the dataset utilized in this study competition hosted by StockX. All sales of Off-White x Nike and Yeezy 350 sneakers from January 9, 2017 (the month Off-White unveiled "The Ten" collection) till now are included in the data set. The data set contains 99,956 total sales, 27,794 sales of Off-White, and 72,162 sales of Yeezy. Only sales in the United States are included in the sample. The author started by cleaning the data, which included things like changing the "\$" from the attribute sales price to integer type data, converting the appropriate object type data to dtype, and eliminating the dollar sign. It is also crucial to rename columns to remove spaces. The author plotted the sneakers to explore the dataset.

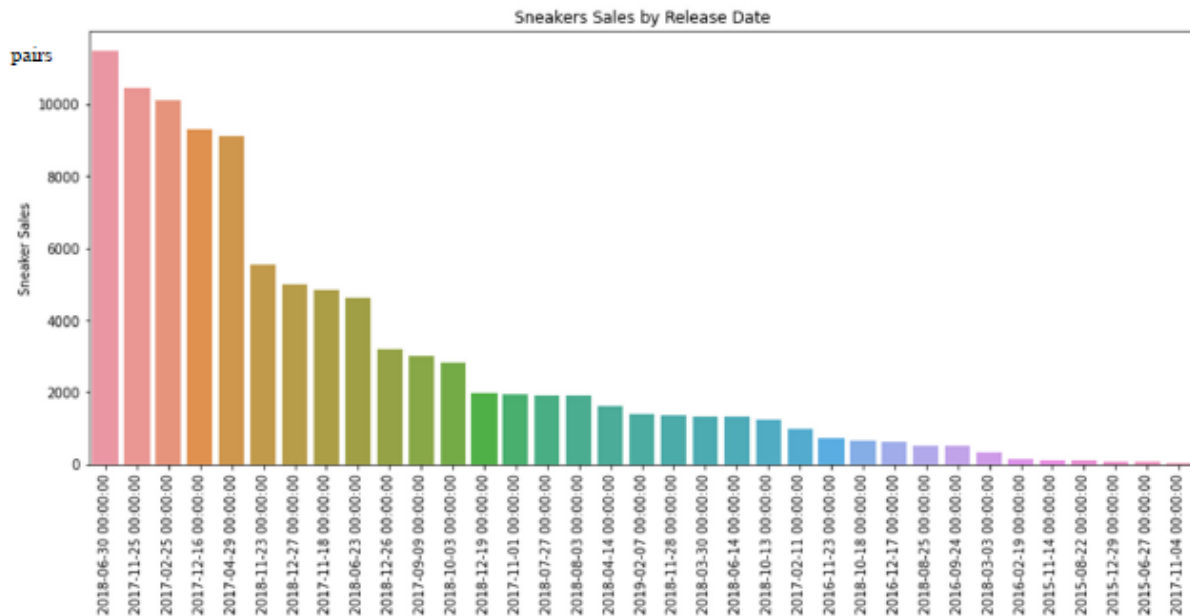


Figure 1. Sneakers sales by release date

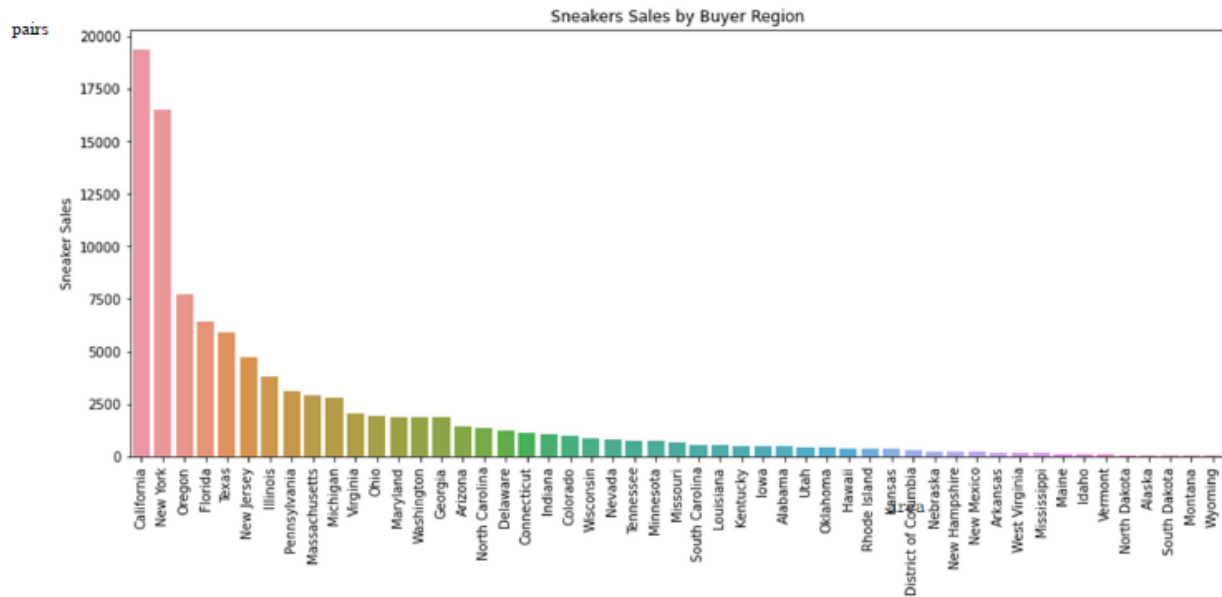


Figure 2. Sneakers sales by buyer region

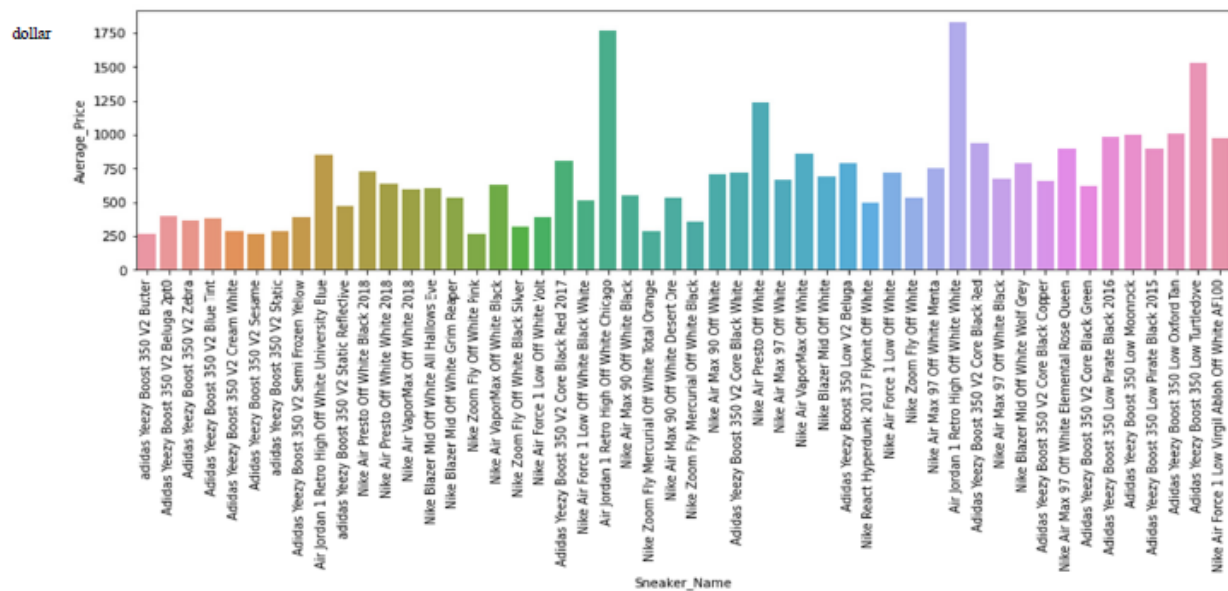


Figure 3. Average price for each sneaker name



Figure 4. Average daily sale price over time

	Coefficient
Sneaker_Name_Nike Blazer Mid Off White All Hallows Eve	1.054632e+09
Sneaker_Name_Nike Blazer Mid Off White Grim Reaper	1.054632e+09
Sneaker_Name_Nike Blazer Mid Off White Wolf Grey	9.870218e+08
Sneaker_Name_Nike Air Max 90 Off White Black	7.333264e+08
Sneaker_Name_Nike Air Max 90 Off White Desert Ore	7.333264e+08
---	---
Sneaker_Name_Adidas Yeezy Boost 350 Low Oxford Tan	-1.587586e+09
Sneaker_Name_Adidas Yeezy Boost 350 Low Moonrock	-1.665599e+09
Sneaker_Name_Nike Air VaporMax Off White	-1.785643e+09
Sneaker_Name_Adidas Yeezy Boost 350 Low Pirate Black 2015	-1.811222e+09
Sneaker_Name_Adidas Yeezy Boost 350 Low Turtledove	-1.908304e+09

107 rows × 1 columns

Figure 5. Coefficient scores of each variable

Next, we'll need to execute assessment metrics and save the forecasts. The author employs MAE to test in order to observe the model's outcome. It is the average mistake that is known as the Mean Absolute Loss (MAE) in regression models. When you add up the absolute values of the target value and forecast value's difference, you obtain the mean absolute error, or MAE. Without taking direction into account, it can only assess the average modulus of the prediction error, which may take values between zero and positive infinity.

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (9)$$

MAE value for this model is: 60.3380912301444.

IV. RESULT EVALUATION

The author used the model's test data [11]. The author used the seaborn function distplot() to create a graph depicting the correlation between y_train and y_train_price.

Figure 6 shows that the majority of the y-testing samples are just equal to 0.01.

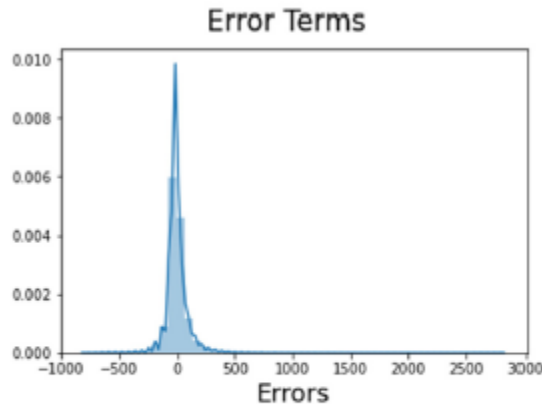


Figure 6. Error terms

The author plotted y_{test} and y_{pred} to understand the spread. X axis is y_{test} , and the y axis is the y_{pred} .

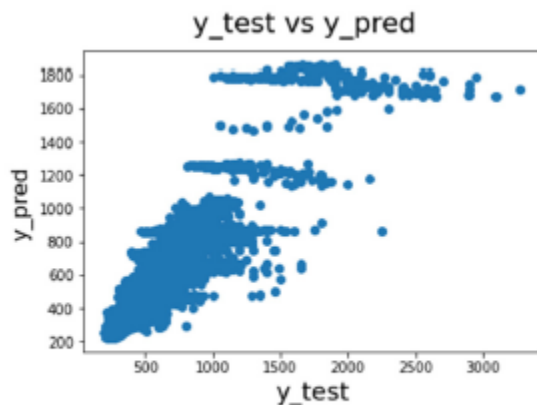


Figure 7. y_{test} -- y_{pred}

DISCUSSION

The simplicity and clarity of linear regression make it a popular choice for use in sales and pricing forecasting models. Not to mention that it is unable to handle nonlinear data, this technique may be too sensitive when dealing with overfitting, outliers, and multicollinearity. Following the evaluation of the data and the training of the machine learning model, the OLS model's accuracy has achieved a certain level after modifications to the model's functioning and the adoption of new testing criteria. The accuracy will naturally vary from run to run due to the fact that the test and training data are not identical. Only a linear regression model is used to forecast shoe prices due to the small number of dimensions in the dataset. The results of the thermodynamic diagram suggest that ordinary least squares (OLS) is the best fit for this dataset. Predicting the overall direction of the sneaker industry using machine learning is fraught with difficulty. Why? Because there are a lot of things that influence the general direction of shoe prices. Take, for instance, the skyrocketing price of comparable autographed brand shoes caused by the retirement of some basketball superstars. Fashion trends are notoriously hard to foresee. Data science and machine learning are the

exclusive lenses through which this study examines and makes predictions about the stock data set. Consequently, inadequate precision is inevitable.

CONCLUSION:

This article achieves the sneaker price prediction provided by the data set by researching and analyzing the stock data competition dataset and then using machine learning models to the changed dataset. For some shoe-selling applications, price prediction is a lifesaver. They are able to provide consumers with the anticipated future cost of a shoe pair excluding the influence of unexpected events. Additionally, this is beneficial for those who purchase shoes. They may also use the shoe price forecast to purchase shoes that will retain their worth. The data set's prediction model may be too susceptible to issues of fitting, outliers, and multicollinearity when using linear regression, despite the fact that this approach is straightforward and easy to comprehend and use. As a result, other models, such as the random forest model, will provide superior results for predictions. Regrettably, the dataset lacks sufficient dimensions to construct an applicable random forest prediction model. Additionally, this study has a few limitations. As an example, there are less dimensions in the dataset. Just two prices—retail and sale—make up the price dimension. If the data set's price property could be extended to include other dimensions, for example, documenting the price trend of shoes over time, the price one month after sale, the price in the previous three days, etc. The prediction accuracy improves as the dataset dimension increases, but the model complexity increases as well.

REFERENCES

1. Leigh Steinberg 2018, The Profitable Hidden Sneaker Market, <https://www.forbes.com/sites/leighsteinberg/2018/09/17/the-profitable-hidden-sneaker-market/?sh=7a39e7aa5925>
2. Altman, N., Krzywinski, M. Simple linear regression. Nat Methods 12,999–1000 (2015). <https://doi.org/10.1038/nmeth.3627>
3. Cohen, J., Cohen P., West, S.G., & Aiken, L.S. Applied multipleregression/correlation analysis for the behavioral sciences. Hillsdale, NJ:Lawrence Erlbaum Associates. 2003.
4. <https://xueshu.baidu.com/usercenter/paper/show?paperid=1j250p00ma0a0ef0hu2v0tm0cv615916>
5. Cohen, J., Cohen P., West, S.G., & Aiken, L.S. Applied multipleregression/correlation analysis for the behavioral sciences. Hillsdale, NJ:Lawrence Erlbaum Associates. 2003.
6. S. Richter-Walsh , Ordinary Least Squares (OLS) Linear Regression in R, 2017 <https://www.r-bloggers.com/2017/07/ordinary-least-squares-ols-linear-regression-in-r/>
7. WuJinMei,LinXiaoDong,LiYongGang,Ch nLiang,JinHanJin,Weightedleastsquaresestimationmethodforsolvingnonlinearmultistructureregression model[J]. Statistics and decision making, 2010, 24:11-13.
8. WangYuan.Secondorderleastsquaresestimationoflinearregressionmodel[D].Beijing Jiaotong University,2016.
9. Ghilani,CharlesD.;Paulr.Wolf,Ph.D.(12June2006).AdjustmentComputations: Spatial Data Analysis.
10. AvijeetBiswal,2021,AnEasyGuidetoStockPricePredictionUsingMachine Learning,
11. <https://www.simplilearn.com/tutorials/machine-learning-tutorial/stock-price-prediction-using-machine-learning>
12. Dita Raditya, Nicholas Erlin P, Ferarida Amanda S, Novita Hanafiah,Predicting Sneaker Resale Prices using Machine Learning,ProcediaComputer Science,Volume 179, 2021, Pages 533-540, ISSN 1877-0509,
13. AMAN KHARWAL,Car Price Prediction with Machine Learning, 2021,<https://thecleverprogrammer.com/2021/08/04/car-price-prediction-with-machine-learning/>