

Machine Learning Algorithms-Based Price Prediction of Sneakers

¹ Inti Vasavi, ² A. N. L Kumar,

¹ MCA Student, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280,

inti.vasavi@gmail.com

² Associate Professor, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280,

Abstract: Sneaker culture is a self-organizing culture spontaneously formed by teenagers. In the Internet and digital age, sports shoes are endowed with appreciation and financial attributes through manual speculation, which has triggered awake of sports shoes speculation with the participation of the whole people. As human group behaviour, the detonating mechanism behind speculative sneakers is worth thinking about. There is a set of logical behaviour chains and practical logic in the behaviour of buying shoes. It is necessary to use threshold theory to reveal the dynamic process of convergence from micro behaviour to macro behaviour. Therefore, machine learning can be used to predict the movement of sneaker prices in the sneaker market. The data set of this study is from stock. In the research, the machine learning method is used to study the price of sports shoes and analyze the data. Four linear regression models were used for prediction. It is found that the OLS regression model's accuracy is high. So, OLS model is the best model for this dataset.

Keywords- Machine learning; linear regression; price prediction; sneakers price

I. INTRODUCTION

The take a look at determined that within the length between 2011 and 2019, the size of the global sports footwear marketplace continued to unfold, and the rate of upward thrust showed an uptrend. In 2018, the market scale of shoes reached 146.5 billion US bucks, and growth of US \$seventy two. Three billion over 2011. For the shoe

market, which has economic attributes, there's less research on the application of machine studying in predicting shoe fees. Therefore, on this study, the precise hassle to be studied is the way to as it should be examine the stock dataset, and what kind of mode list most appropriate for the statistics set based totally at the records evaluation. The prediction of shoe fees

could be very conducive to the manager of shoe expenses by means of the shoe platform and the evaluation of the faculty marketplace by using shoe dealers, to attain more price retaining commodities.

II METHODOLOGY

A. Dataset Analysis and Data visualization

All Off-White x Nike and Yeezy 350 income from January nine, 2017 (the month that Off-White first debuted “The Ten” series) to the present. There are 99,956 total incomes within the data set, 27,794 Off-White sales, and 72,162 Yeezy sales. The sample includes U.S. Income only. Firstly, the writer wiped clean the statistics, such as changing the relevant item kind statistics to kind, removing the "\$" within the attribute sales fee and changing it to integer kind statistics. Renaming columns tiered of areas is likewise crucial too.

The result of this wide variety is the mean rectangular error, and the denominator is the sample size minus the quantity of parameters to be anticipated. The fashionable error of these parameter estimates is shown as follows:

Sales via sneaker call, sneaker sales by using show length. Then the fashion

among sneaker call and sale price turned into analyzed. The sale charges right here is the common sale price for every sneaker name. After these, the author created a visualization of the average sale rate through the years [8].

Sales and date is that the sooner the time, the less the sneakers income.

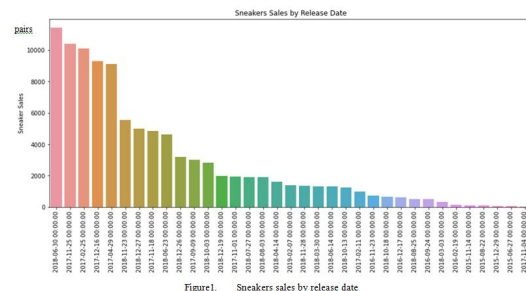


Figure1. Sneakers sales by release date

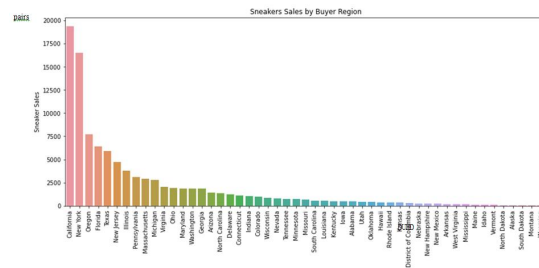


Figure2. Sneakers sales by buyer region

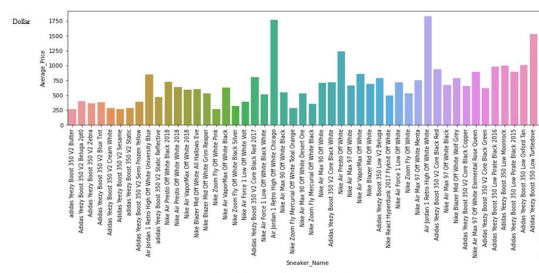


Figure3 Average price for each sneaker name



Figure4 Average daily sale price overtime

Modelling

The attribute to be predicted is the sale price. Firstly, the author split the training

set and the test set. The author set the parameter: test size=0.2. The training set has 80% of the Regression model's intercept. The result is 1274056226783.0996. Fourthly, the coefficient scores of each variable need to be shown.

	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 x 1 columns

Figure5. Coefficient scores of each variable

The next is to keep predictions and run evaluation metrics. To see the result of the model, the writer uses MAE to check. The Mean absolute loss (MAE) in the regression model is the average blunders. MAE is the sum of the absolute values of the distinction among the target value and the predicted price. It handiest can measure the average modulus of the prediction mistakes without thinking about the path, and the cost variety is zero to positive infinity.

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

MAE value for these models: 60.3380912301444.

III RESULT EVALUATION

The author took the test data on the model [11]. The author plotted the relationship between y_{train} and y_{train_price} , and used the seaborn's function `despot()` to draw the picture.

Infigure6, it is can be seen that most of the testing samples of y just equals 0.01.

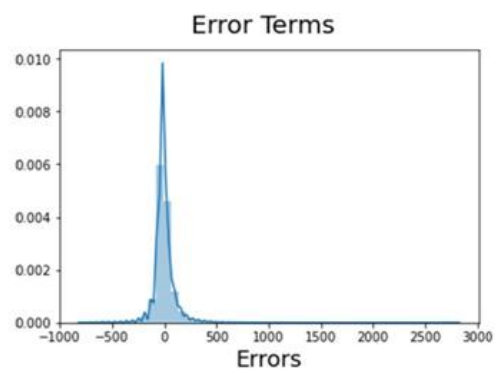


Figure6. Error terms

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} .

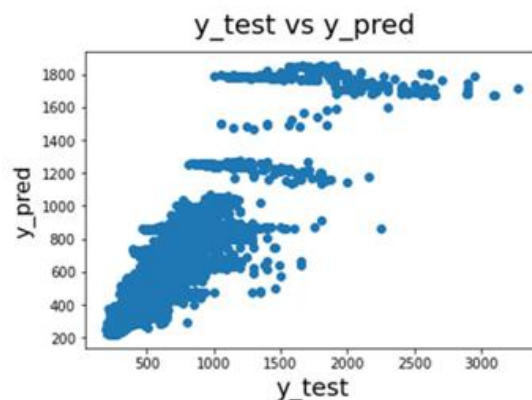


Figure7. y_{test} - y_{pred}

IV. DISCUSSION

Linear regression is broadly utilized in fee and income forecasting models because it

is easy to understand and explain. However, this method can be too touchy when managing over fitting, outliers and multicollinearity, not to mention it cannot capture nonlinear statistics. After the data analysis effects and machine learning education, the accuracy of the OLS model has reached a positive degree after changing the operation of the version and different take a look at requirements. Of direction, because the take a look at and education records divided by means of every run are exclusive, there will be a positive deviation inside the accuracy. Because the facts set have few dimensions, only linear regression model is used to predict the price of sneakers. After drawing the thermodynamic diagram, its miles discovered that OLS is probably the most suitable for this records set. It isn't always very accurate to be expecting the trend and fashion of the sneaker market based on device learning technology from a macro factor of view. Because many factors affecting the charge fashion of the shoe market. For example, because of the retirement of a few basketball stars, the charge of the corresponding signed emblem footwear will even upward thrust sharply. The fashion of favour is sort of unpredictable. This research only analyzes and predicts the records set given with the aid of stock from the angle of facts

technology and machine studying. Therefore, there will unavoidably be inadequate accuracy.

V CONCLUSION

Through the studies and analysis of the information set of stock records opposition and the modelling of gadget studying for the changed facts set, this paper realizes the prediction of sneaker fee given by using the records set. The prediction of shoe rate may be very practical in some shoe selling applications. They can give users the destiny forecast rate of a couple of footwear without the impact of emergencies. At the equal time, this is also favourable for shoe buyers. The prediction of shoe charge also can help them purchase a pair of shoes with value maintenance. For the prediction version mounted by the statistics set, although linear regression is a simple and clean technique to recognize and put in force, this method may be too sensitive when dealing with fitting, outliers and multicollinearity. Therefore, using different fashions which includes the random woodland version may have a better prediction effect. However, it's far a pity that the facts set have too few dimensions to set up an applicable random woodland prediction model. There are also

some deficiencies in this study. For instance, the records set have fewer dimensions. For the charge size, there are best the sale price and retail charge. If the rate attribute of the information set may be elevated to more dimensions, consisting of recording the fee trend of a pair of shoes consistent with time, the price one month after sale, the fee in the last three days, and so forth. With the increase of dataset dimension, the extra complicated the version might be, but the better the accuracy of prediction will be.

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 multiple regression/correlation analysis for the behavioural sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates. 2003.
4. Cohen, J., Cohen P., West, S.G., & Aiken, L.S. *Applied multiple regression/correlation analysis for the behavioural sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates. 2003.
5. Prasadu Peddi (2015) "A review of the academic achievement of students utilising large-scale data analysis", *ISSN: 2057-5688, Vol 7, Issue 1, pp: 28-35*
6. Prasadu Peddi (2015) "A machine learning method intended to predict a student's academic achievement", *ISSN: 2366-1313, Vol 1, issue 2, pp:23-37*.
7. Prasadu Peddi (2023). *AI-Driven Multi-Factor Authentication and Dynamic Trust Management for Securing Massive Machine Type Communication in 6G Networks*. *International Journal of Intelligent Systems and Applications in Engineering*, 12(1s), 361–374.