IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Sneaker Price Prediction Using Machine Learning: A Systematic Review

Asso.Prof. Archana Agarwal
Department of Computer Science and Engineering
Inderprastha Engineering College,Sahibabad,India

Gunesh Munjal
Department of Computer Science and Engineering
Inderprastha Engineering College,Sahibabad,India

Naman Dubey
Department of Computer Science and Engineering
Inderprastha Engineering College,Sahibabad,India

Divyansh Sharma
Department of Computer Science and Engineering
Inderprastha Engineering College,Sahibabad,India

Joel George
Department of Computer Science and Engineering
Inderprastha Engineering College,Sahibabad,India

Abstract- The sneaker market has evolved into a multi-billion-dollar industry, with sneaker prices often exhibiting significant volatility driven by factors such as brand, exclusivity, release frequency, and market trends. Predicting sneaker prices has become a crucial task for both consumers and businesses, as it can influence purchasing decisions, reselling strategies, and inventory management. Machine learning (ML) techniques have shown considerable promise in modeling and forecasting sneaker prices by leveraging historical data, product attributes, and external factors. This review paper explores the application of various machine learning methods

Keywords- Machine learning.

I. INTRODUCTION

The sneaker market has transformed into a global phenomenon, driven by consumer demand, limited-edition releases, and the rise of sneaker culture. What was once a niche segment in the footwear industry has now become a dynamic and highly competitive market, with sneaker prices often fluctuating significantly based on factors such as brand, rarity, popularity, and market trends. In particular, the secondary sneaker market, where resellers and collectors trade shoes at a premium, has seen tremendous growth, with some sneakers selling for several times their original retail price. This rapid price escalation has sparked interest in predictive models that can anticipate sneaker prices, aiding both consumers and businesses in making informed decisions.

In recent years, machine learning (ML) techniques have emerged as a powerful tool for predicting sneaker prices, leveraging vast amounts of data from various sources. These techniques, ranging from traditional regression models to sophisticated deep learning algorithms, are employed to analyze and forecast sneaker price trends based on historical sales data, release information, product attributes, and external factors such as celebrity endorsements and social media buzz. The ability to accurately predict sneaker prices offers valuable benefits for sneakerheads, resellers, retailers, and investors alike, as it allows them to anticipate market movements and optimize purchasing and selling strategies.

However, predicting sneaker prices presents a unique set of challenges. The market is characterized by its inherent volatility, driven by both predictable and unpredictable factors, such as sudden shifts in consumer preferences or the announcement of surprise collaborations. Additionally,

the sneaker market's complex ecosystem, involving different distribution channels, global supply chains, and fluctuating demand, makes it difficult to model accurately using conventional methods. Furthermore, the large volume of unstructured data from social media, online forums, and influencer marketing complicates the process of feature engineering and data preprocessing.

This review paper aims to provide a comprehensive overview of the current state of sneaker price prediction using machine learning. We will explore the various machine learning techniques employed in this domain, the datasets and features commonly used, and the performance metrics applied to evaluate model effectiveness. In doing so, we highlight the strengths and limitations of existing approaches and suggest potential avenues for future research, including the incorporation of real-time data and hybrid modeling techniques that combine multiple machine learning paradigms. By understanding the progress and challenges in this field, we aim to contribute to the ongoing development of more accurate and reliable predictive models for sneaker pricing.

II. LITERATURE SURVEY

[1] Predicting Sneaker Resale Prices Using Machine Learning

Dita Raditya, Nicholas Erlin P, Ferarida Amanda S, Novita Hanafiah

The unpredictability of sneaker resale prices presents a challenge and a potential business opportunity, motivating this study to explore suitable machine learning algorithms for accurate price forecasting.

The study examines the effectiveness of two machine learning models—linear regression and random forest—in predicting sneaker resale prices based on historical sales data obtained from StockX, a prominent online marketplace. Various machine learning models are employed across industries to predict prices, often relying on historical data and influential features. Past research shows that models like linear regression, support vector machines (SVM), and random forest have been successfully applied in areas like car price prediction and housing cost estimation. In particular, linear regression is commonly used due to its simplicity and interpretability, but it has limitations, including sensitivity to outliers and difficulty capturing non-linear patterns. Random forest, on the other hand, has proven useful for its high accuracy and robustness to noise, although it requires more computational resources.

The authors highlight that the linear regression model operates by fitting a linear equation to data, offering a straightforward and interpretable relationship between independent and dependent variables. Despite its popularity, linear regression is susceptible to several limitations, including overfitting, sensitivity to outliers, and struggles with multicollinearity. These limitations necessitate extensive data pre-processing to ensure model reliability, often making it challenging to achieve high accuracy in complex datasets like sneaker resale prices, which may contain non-linear patterns.

Random forest, a more complex ensemble algorithm, is noted for its superior ability to capture complex relationships in data through the aggregation of multiple decision trees. This approach allows random forest models to offer high predictive accuracy and resilience against overfitting, making them a suitable choice for applications involving substantial variability. However, the method's computational demands and need for longer training times are recognized as potential drawbacks, though they are often offset by improved performance in handling noisy or outlier-prone datasets.

The study's methodology involves a detailed feature engineering process, where variables like sneaker brand, release date, and past sales prices are incorporated into both models to enhance predictive accuracy. The dataset is split into training and test sets, with 80% of the data used for training and 20% for testing. Initial model evaluations use R-squared and Mean Squared Error (MSE) as performance metrics. Linear regression initially shows a slight advantage in R-squared values. However, after applying cross-validation, random forest outperforms linear regression, indicating that random forest's ensemble structure provides greater stability against overfitting.

The study concludes that while both linear regression and random forest are viable for sneaker resale price prediction, random forest ultimately performs better when cross-validation is applied, enhancing its R-squared score and reducing MSE. This result highlights random forest's suitability for complex data with minimal pre-processing, making it advantageous for real-world applications where time and computational resources may be limited. The authors suggest that future research could incorporate social media data to account for unpredictable price spikes due to events or endorsements, potentially improving the model's adaptability to sudden market shifts.

Overall, this review underscores the relevance of machine learning in pricing models for the sneaker resale market, demonstrating that random forest, with its flexibility and robustness, offers a promising solution in comparison to simpler linear models.

[2]From Local to Global in the Sneaker Universe: a Data Science approach

Luciano Perdomo, Leo Ordinez

The document titled From Global to Local in the Sneakers Universe: A Data Science Approach investigates the application of machine learning to predict sneaker prices in Argentina, particularly within the context of the COVID-19 pandemic, which significantly boosted e-commerce in the region. The study's aim is to provide local businesses with insights to enhance decision-making by predicting sneaker prices and analyzing competition with national and global companies. This study also emphasizes the accessibility of data science methodologies for smaller companies looking to compete in the online retail market,

particularly through cost-effective tools such as web scraping and machine learning models.

The research methodology begins with data collection through web scraping from several Argentine e-commerce websites, categorizing retailers into global, national, regional, and local levels. The authors gathered data on brands, prices, discounts, and gender-specific categories, structuring the dataset to allow comparison across different market levels. They utilized tools like ParseHub and BeautifulSoup for data extraction, followed by exploratory and predictive analyses on variables such as brand, company, and gender. This categorization into multiple scales highlights the study's multiscalar approach, which is rooted in the understanding that e-commerce inherently provides broad customer access.

Three experiments were conducted to evaluate various machine learning models, assessing their effectiveness for price prediction. Experiment 1 compares linear models—linear regression, ridge regression, and stochastic gradient descent (SGD) regression—finding that linear regression with polynomial features yields the highest R-squared values. Experiment 2 assesses decision tree-based models—random forest, XGBoost, and decision tree regressors—showing that XGBoost and random forest perform comparably, though random forest achieves slightly better results after hyperparameter tuning. Experiment 3 tests support vector machine (SVM) and LightGBM models, with LightGBM emerging as the best-performing model in terms of R-squared accuracy. For each experiment, the data is divided into training and testing subsets, and performance is measured across companies.

The analysis indicates that global companies achieve more accurate price predictions compared to local companies, suggesting that they benefit from greater market stability and larger datasets. Stockcenter, a global retailer, consistently performs the best across experiments, while regional and local companies like Sporting and Ferreira show lower predictive accuracy. Notably, company-specific models show that price prediction accuracy improves with company scale, but Quonam, a local company, performs surprisingly well, likely due to effective outlier handling in data preprocessing.

The study concludes that while global companies have an advantage in predictive accuracy, local retailers can still leverage machine learning models to enhance their competitive position in e-commerce. Incorporating more data, such as sales volume or additional sneaker attributes, could refine predictions further. Future work includes gathering data weekly to capture price trends and potentially using a data warehouse to expand data sources and analysis frequency.

This research underscores the viability of data science for improving price prediction models across varying business scales. By using adaptable, affordable methodologies, the study provides a blueprint for small and medium businesses aiming to thrive in the e-commerce market.

III. CONCLUSION

In this review, we have explored the growing field of sneaker price prediction using machine learning, highlighting its potential to revolutionize decision-making in the sneaker market. The ability to forecast sneaker prices accurately offers significant value to consumers, resellers, retailers, and brands by providing insights into market trends, optimizing sales strategies, and guiding investment decisions. Through the application of machine learning techniques such as regression models, decision trees, deep learning, and ensemble methods, researchers have demonstrated the effectiveness of data-driven approaches in understanding the complex dynamics of sneaker pricing.

Despite significant advances, challenges remain in creating robust predictive models. The sneaker market's inherent volatility, combined with the dynamic nature of consumer preferences, external factors like social media influence, and limited availability of high-quality data, makes price prediction a difficult task. Furthermore, the use of unstructured data (e.g., online reviews, influencer posts) and the variability across different markets or regions adds another layer of complexity to the models. Overfitting, data sparsity, and the continual evolution of trends also present hurdles in developing models that maintain accuracy over time.

Nevertheless, the intersection of machine learning with sneaker price prediction holds great promise. Future research could focus on improving data collection techniques, especially through the integration of real-time information and unstructured data sources. Hybrid models, which combine various machine learning approaches or incorporate domain-specific expertise, could further enhance predictive accuracy. Additionally, the exploration of transfer learning, reinforcement learning, and multi-modal data fusion may offer new avenues for handling the challenges posed by this complex, ever-changing market.

In conclusion, while significant strides have been made in sneaker price prediction using machine learning, continued innovation and adaptation are essential to keep pace with the fast-evolving landscape of sneaker culture and commerce. As these models become more sophisticated, they will likely play an increasingly vital role in shaping consumer behaviour, guiding pricing strategies, and creating new opportunities within the sneaker industry.

IV. FUTURE SCOPE

The field of sneaker price prediction using machine learning has made significant strides, yet there remains considerable room for innovation and improvement. As the sneaker market continues to evolve, so too must the methodologies and models that seek to predict its pricing trends. The following areas represent key opportunities for future research and development in this domain:

Incorporation of Real Time and Dynamic Data : One of the most promising directions for future research is the integration of real-time data into prediction models. Sneakers often experience price fluctuations driven by short-term factors such as sudden influencer endorsements, viral trends, and social media buzz. Incorporating real-time data from social platforms, auction sites, and news sources can help capture these dynamic shifts and allow models to respond more rapidly to market changes..

Reinforcement Learning for Dynamic Pricing: Reinforcement learning (RL) offers a promising avenue for real-time dynamic pricing strategies. By simulating various pricing scenarios and adapting to changes in consumer demand, RL models can optimize sneaker prices over time. This could lead to more effective pricing strategies for resellers, retailers, and brands, allowing for price adjustments that reflect supply-demand equilibrium and maximize profits.

Explainability and Transparency of Model: As machine learning models become more complex, ensuring their interpretability and transparency will be critical, especially in markets with significant economic impact. Future research should focus on developing explainable AI (XAI) frameworks that can clarify how different factors (e.g., release dates, brand collaborations, and scarcity) influence price predictions.

Integration of Supply Chain and Inventory Data: Price predictions could be further refined by integrating supply chain and inventory data, which can offer insights into scarcity and availability, both of which directly influence sneaker prices. Modeling the relationship between inventory levels, release strategies, and price dynamics could enable a more accurate assessment of potential price volatility based on production cycles, stock levels, and distribution delays.

Transfer Learning and Domain Adaption: As sneaker markets are highly region-specific and may vary based on geographical location, using transfer learning techniques to adapt models to different markets could improve prediction accuracy. Transfer learning could allow models trained on data from one region to be fine-tuned for other regions, especially where sneaker culture or demand patterns differ.

V. REFERENCES

[1] Dita Raditya, Nicholas Erlin P, Ferarida Amanda S and Novita Hanafiah"Predicting Sneaker Resale Prices Using Machine Learning" 5th International Conference on Computer Science and Computational Intelligence 2020.

[2] Luciano Perdomo and Leo Ordinez"From Global to Local in the Sneaker Universe : A Data Science Approach" Laboratorio de Investigaci´on en Inform´atica (LINVI).

[3] WeiJia Yan"2022 International Conference on Machine Learning and Knowledge Engineering (MLKE).

[4] Kiattisak Rattanaporn"Shoes Price Dataset"Kaggle.com, 2023.

[5] Duyku Guzel, M.S. Bostanzi and Askerzade"A Machine Learning Based Approach for Price Estimation" In Proceedings of IAM, V.11, N.1, 2022, 50-61.

[6] Zhang, Tony"Predicting Sneaker Resell with Deep Learning"September 22, 2020.

[7] Statista "Sneaker-Worldwide | Statista MarketForecast" Statista 2024.

[8] Sateesh Vishnu"Hyper Parameter Tuning(GridSearch CV or RandomizedSearchCV)" Analytics Vishya, April 18, 2021.

[9] Garcia, Jolie"Sneaker Prices Reach Highest Levels This Decade-FDRA"FDRA, April 13, 2023.

[10] Akshar Sinha"Predicting Shoe Prices Using Machine Learning Algorithms"August 6, 2023.