iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.61754

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Dynamic Pricing for Anomaly Detection in Online Market Places

Krishnapriya C[1], Goury Priya Sajeev[2], Thejas R[3], Sreelakshmi K B[4], Leya Elizabeth Sunny[5], Joby Anu Mathew[6]
Department of Computer Science and Engineering Mar Athanasius College of Engineering, Kothamangalam, Kerala

Abstract: ***Online marketplaces have become integral parts of e-commerce, providing convenient platforms for buying and selling goods. Detecting fraudulent listings is crucial for maintaining trust and integrity within the marketplace. This project addresses this challenge by proposing a method for fraud detection based on predicting listing prices and identifying discrepancies between predicted and listed prices and also updating prices based on current market conditions. By using machine learning techniques to predict listing prices based on working features, customer demand and market status, the aim is to uncover potentially fraudulent listings where the listed price significantly deviates from the predicted value. The project utilizes a dataset containing information about listings on the online marketplace, including features such as product category, description, and location, along with the listed prices. Data preprocessing techniques are applied to clean and prepare the dataset for analysis. Machine learning algorithms, including regression models and natural language processing techniques for textual data, are employed to predict listing prices based on their features. This also includes some methods of feature engineering. In conclusion, this project presents a novel approach to fraud detection in online marketplaces by leveraging machine learning techniques for dynamic price prediction. By identifying discrepancies between predicted and listed prices, the developed model effectively detects potentially fraudulent listings. The integration of this model into the marketplace platform enhances security and reliability, fostering a safer and more trustworthy environment for buyers and sellers. Moving forward, further refinement and optimization of the model can lead to even greater accuracy and effectiveness in fraud detection.***

Index Terms: ***component, formatting, style, styling, insert***

I. INTRODUCTION

In the vast expanse of online commerce, where convenience and accessibility thrive, lurk challenges that threaten consumer trust and market fairness. Among these challenges, two promi-nent issues stand out: unreliable price listings and fraudulent activities. While they may not always grab headlines, these issues quietly undermine the integrity of e-commerce plat- forms, prompting concerns among consumers and industry stakeholders alike.

The landscape of online shopping has expanded exponen- tially in recent years, offering consumers unparalleled access to a global marketplace. However, with this growth comes the inevitable presence of unreliable price listings. In the dynamic realm of e-commerce, prices fluctuate constantly, influenced by various factors such as demand, supply, and pricing al- gorithms. Yet, discrepancies between advertised prices and actual costs often occur, whether due to technical glitches or unintentional errors. Such discrepancies can erode consumer trust and highlight the need for greater transparency in pricing practices.

In addition to unreliable price listings, the specter of fraud looms over e-commerce platforms. Cybercriminals employ various tactics, from identity theft to payment fraud, to exploit vulnerabilities and defraud unsuspecting consumers. Despite efforts to combat these illicit activities, the evolving nature of cybercrime presents ongoing challenges for platform operators and regulatory authorities alike.

While these issues may not be as sensational as they sound, they nevertheless warrant attention and concerted efforts to address them. By fostering transparency, implementing ro- bust security measures, and enhancing consumer awareness, stakeholders can work together to create a more resilient and trustworthy e-commerce environment. Through collaborative efforts, we can mitigate the risks posed by unreliable price listings and fraud, ensuring that the promise of online shopping remains both convenient and reliable for all consumers.

II. RELATED WORKS

A. Dynamic Pricing Algorithm

Dynamic pricing, also known as surge pricing or demand pricing, is a pricing strategy in which businesses adjust the prices of their products or services based on various factors such as demand, time of day, competitor pricing, and inven- tory levels.

This strategy is prevalent in industries such as transportation (e.g., ride-sharing services like Uber and Lyft), hospitality (e.g., hotel bookings), e-commerce, and entertain- ment (e.g., ticket sales for concerts and sporting events).
The main goal of dynamic pricing is to optimize revenueby aligning prices with market demand and maximizing prof- itability. By leveraging real-time data and sophisticated algo- rithms, businesses can dynamically adjust prices to reflect fluc- tuations in demand and supply conditions. For example, during peak hours or high-demand periods, prices may increase to capitalize on consumer willingness to pay, while during off- peak times, prices may decrease to stimulate demand and fill excess capacity.

B. Web Scraping

Web scraping is the process of extracting data from web- sites. It involves accessing the HTML or other structured data of a webpage and parsing it to extract the desired information automatically. Web scraping can be performed using various tools and programming languages, such as Python, JavaScript,or specialized web scraping libraries like BeautifulSoup and Scrapy.

C. Ensemble Learning

Ensemble learning involves combining multiple models to make predictions. Theory and Practice: Explanation of ensemble learning principles, including bagging, boosting, and stacking.Theoretical foundations of Random Forest, its decision-making process, and advantages over individual de- cision trees.Practical considerations in implementing and tun- ing Random Forest for data leakage detection and malware scanning tasks.

D. Data Preprocessing

Data preprocessing is an essential phase in the data analysisprocess, serving as the foundation for extracting meaningful insights and building reliable models. It encompasses a series of steps aimed at cleaning, transforming, and structuring raw data to make it suitable for analysis.In this phase, data issubjected to several operations, including handling missing values, addressing duplicates, encoding categorical variables, scaling numerical features, and reducing dimensionality. Han- dling missing values involves strategies such as imputation or deletion to ensure data completeness.

E. Natural Language Processing

Natural Language Processing (NLP) is a branch of ar- tificial intelligence that focuses on the interaction between computers and human languages. It enables computers to understand, interpret, and generate human language data in a meaningful way. In Natural Language Processing (NLP), tokenization and lemmatization are essential techniques for processing and analyzing textual data. Tokenization breaks down a piece of text into individual units, such as words or sentences, making it easier for computers to understandand manipulate. This process serves as the foundation forvarious NLP tasks, including sentiment analysis, named entityrecognition, and machine translation. Lemmatization , aims to reduce words to their base or root form, known as the lemma. This helps standardize words and reduces the complexity of the vocabulary, improving the accuracy of NLP models. Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) technique used to analyze and extract subjective information from textual data.

F. Evaluation Metrics

Evaluation metrics are used to assess the performance of machine learning models.Metrics like Mean Squared error, Root mean squared Error can be used in evaluating a model performance. Moreover, model performance can be visualisedand then evaluated

III. PROPOSED MODEL

The primary aim of this project is to mitigate fraud inonline marketplaces by employing dynamic pricing algorithmsand predictive models.Regression models are trained on the processed data to predict the expected/fair price for various products like cars, clothing, shoes, appliances, etc., based on their descriptions and attributes. The predicted prices fromthe regression models are compared against the listed prices on the online marketplaces. Significant deviations between the predicted and listed prices could indicate potential fraud or price manipulation. This fraud detection capability, along with the pricing predictions, can be integrated into the online marketplace website to alert users about suspicious listings andprovide them with estimated fair prices for better decision- making. The system also supports typical e-commerce func- tionalities like product listings, purchase carts, etc., likely utilizing the pricing predictions. The system leverages data- driven pricing models to detect fraud and enable fair pricing transparency for users in online marketplaces.

The project aims to provide a comprehensive system that integrates data ingestion, preprocessing, storage, natural language processing, and machine learning techniques to detect potential fraud and enable fair pricing in online marketplaces.

IV. DYNAMIC PRICE PREDICTION

This project employs a dynamic pricing method that in- tegrates Natural Language Processing (NLP) techniques and multiple machine learning (ML) models to generate precise and equitable pricing predictions for products listed on online marketplaces. The project Uses ensemble methods to train product data in different machine learning models, refered to as stacking or stacked generalisation to get an accurate prediction.Fraud detection is employed by integration of the predicted price and the results of sentiment analysis of productdescription using NLP to detct the level of aunthenticity of theproduct and its seller.

Natural Language Processing (NLP): Utilizes NLP tech-niques such as tokenization, lemmatization, and named entity recognition (NER) to process textual product descriptions. These techniques extract pertinent features like product type, brand, model, condition, and age from the text. Feature En- gineering: Combines NLP-extracted features with structured data like product categories, marketplace details, and histori- cal pricing data. Feature engineering crafts a comprehensive feature set capturing factors influencing product pricing effec- tively. Ensemble Modeling: Adopts an ensemble approach by training multiple ML models like linear regression, decision trees, random forests, gradient boosting, and neural networks on the feature-engineered dataset. This strategy enhances prediction accuracy and robustness. Model Stacking: Stacks predictions from individual ML models using techniques like averaging or meta-modeling to capitalize on each model's strengths and mitigate weaknesses. This enhances the relia- bility of pricing predictions. Dynamic Updating: Periodically re-executes feature engineering and ensemble modeling as new product listings, descriptions, and pricing data become available. This ensures that pricing predictions reflect current market conditions. Anomaly Detection: Compares ensemble model predictions with actual listed prices to detect sig- nificant deviations, indicating potential anomalies or pric- ing manipulation. Pricing Recommendations: Provides pricing recommendations for products with suspected fraudulent or manipulated pricing based on ensemble model predictions.These recommendations empower marketplace administrators or users to make informed decisions. Continuous Monitoring: Operates in a continuous loop, constantly monitoring new list-ings, updating ML models, and providing pricing predictions and recommendations. This ensures adaptability to evolving market conditions and effectiveness in detecting anomalies. By integrating NLP techniques, ensemble modeling, dynamic updating, anomaly detection, and continuous monitoring, this dynamic pricing method promotes pricing transparency, de- tects fraudulent activities, and fosters trust and fairness inonline marketplaces.

.

A. Data Collection

Data collection encompasses acquiring, compiling, and preprocessing raw data from diverse sources for analysis. It involves identifying pertinent sources, accessing andextracting data in a structured manner, and meticulously cleaning and transforming it to ensure accuracy and consistency. Additionally, data collection necessitates adherence to privacy regulations and documentation of the process for transparency and reproducibility. This iterativeprocess lays the groundwork for informed decision-making, research insights, and strategic planning across variousdomains, from scientific research to business analytics.

B. Data Preprocesing

Data preprocessing is a vital stage in the data analysis pipeline where raw data undergoes cleaning, transformation, and organization to enhance its quality and suitability for analysis. This process includes handling missing values, removing duplicates, encoding categorical variables, scaling numerical features, and performing feature engineering to create new variables or derive insights. Data preprocessing aims to standardize and normalize the data, making it more conducive to modeling and analysis. Additionally, preprocessing involves splitting the data into training, validation, and testing sets for model development and evaluation.

C. Data Evaluation

Data evaluation involves assessing the quality, relevance, and validity of collected data to determine its suitability for analysis. This process entails examining the data for accuracy, completeness, and consistency while ensuring it aligns with project objectives. Additionally, data evaluation involves checking the validity of the data against external sources or expert knowledge and assessing its format, structure, and accessibility for analysis. Evaluators also consider potential biases and ethical implications, ensuring fairness and compliance with regulations. By rigorously evaluating the data, analysts can identify any limitations or biases that may affect analysis outcomes and make informed decisions based on trustworthy data..

D. Feature Extraction

Using Random Forest, pattern detection entails applying an ensemble learning method to find connections and patterns in a dataset. Using random subsets of the data and features, Random Forest builds several decision trees, then aggregates the predictions to increase precision and resilience. Complex patterns, nonlinear relationships, and variable interactions can all be found using this method. Random Forest is aneffective tool for tasks involving classification, regression,and anomaly detection in a variety of industries, including marketing, finance, and healthcare. It does this by examining the decision paths and feature importance of the trees, which allows it to reveal important patterns and insights in the data.

E. Feature Transformation

Feature Transformation involves converting extracted fea- tures into a format that is suitable for machine learning algo- rithms. This process aims to improve computational efficiencyor enhance model performance by reducing the dimensionalityof the feature space or transforming the features into a more informative representation. Vectorization is a common tech- nique used to represent textual or categorical features as nu- merical vectors, such as one-hot encoding or term frequency- inverse document frequency (TF-IDF) encoding. This allows algorithms to process the data efficiently and uncover patternsin the text.

F. Splitting the dataset

Before training any model, it's essential to split the dataset into training, validation, and testing sets. The training set is used to train the model, the validation set helps tune hy- perparameters and assess model performance during training, and the testing set evaluates the final model's performance on unseen data.

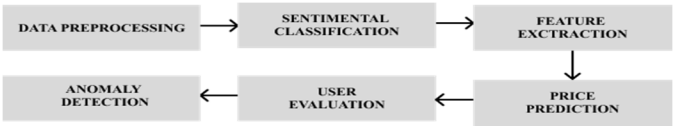


Fig. 1. Implementation Diagram for Dynamic Pricing Algorithm

G. Model Selection

Choosing the appropriate model architecture is crucial for achieving accurate predictions. Depending on the nature of theproblem and the characteristics of the dataset, various models such as Random Forest, Support Vector Machines, or Neural Networks may be considered.

H. Training the model

Once the model is selected, it is trained using the training dataset. During training, the model learns the underlying patterns and relationships in the data through an optimization process that minimizes a chosen loss function.

I. Prediction and Evaluation

After training, the model is used to make predictions onthe validation or testing dataset. The predictions are then evaluated using appropriate evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), or Ac- curacy, depending on the nature of the problem. This step provides insights into the model's performance and its ability to generalize to unseen data.

V. FRAUD DETECTION

In this project fraud detection is latter to price prediction, with analysing simple standard deviation statistic of the product dataset. Fraud detection is employed using the sentimental analysis of the product description and classifyingit as a good or bad feedback.

A. Data Collection:

webscraping Data collection is a foundational step in build- ing a machine learning-based malware detection system. It involves curating a comprehensive and representative dataset comprising both malware and benign files. This dataset should mirror the diversity of potential threats in the real-world environment where the model will be deployed. Ensuring that the dataset covers a wide range of malware types and includessamples relevant to the specific use case is critical. A well- annotated dataset provides the foundation for training a robustand effective malware detection model.

B. Data Preprocessing:

Once the dataset is assembled, the next step is data pre- processing. This involves cleaning and organizing the datato make it suitable for machine learning algorithms. Tasks include handling missing values, scaling numerical features toa standardized range, and encoding categorical variables. The objective is to create a coherent and standardized dataset that can be effectively utilized for training and testing the machine learning model. Proper data preprocessing is essential for the model to learn meaningful patterns and relationships withinthe data.

Fig. 2. Pricing Evaluation : Comparing predicted and actual prices

C. Feature Extraction:

Feature extraction is a crucial phase where relevant charac- teristics are identified and extracted from the dataset. In the context of malware detection, these features could encompass file size, API calls, system calls, or opcode sequences. The selection of features significantly influences the model's abilityto discriminate between malicious and benign files. Thoughtful consideration of which features to include and how to repre- sent them is paramount to the success of the detection system.

D. Price Prediction

Machine learning techniques are commonly used for price prediction, leveraging historical data and relevant features to make predictions about future prices. Some common ap- proaches include: Modeling the relationship between predictorvariables (such as market indicators, economic factors, or product attributes) and the target variable (price) using a regression technique like linear regression. Employing su- pervised learning algorithms such as decision trees, random forests, gradient boosting, or neural networks to learn patterns and relationships in the data and make predictions about futureprices.

E. Statistical Classification

Price is then classified as reasonable or not by using simple standard deviation statistic with a threshold of 20

F. Sentimental Analysis

Sentimental analyisis of the product description analysiscan be performed using various methods, including rule- based approaches and machine learning algorithms. These methods analyze the textual content, identifying sentiment- bearing words or phrases, as well as contextual cues and linguistic patterns that convey sentiment.

G. Classification and Evaluation

Classification analysis feeds into fraud detection and the product is then classified as fraud or not, with listed prices being specified as reasonable or not.

Fig. 3. Deviation of predicted values from current value

VI. RESULT

The project developed a sophisticated dynamic pricing and anomaly detection system tailored specifically for online marketplaces, significantly bolstering security and reliability. Leveraging advanced machine learning techniques, the system accurately predicted listing prices by analyzing a multitudeof factors, including product features, market demand, and prevailing conditions.

This predictive capability enabled the system to swiftly identify potentially fraudulent listings by detecting significant disparities between projected and listed prices. By integrating seamlessly into existing marketplace infrastructure, the system offers real-time monitoring and alerts, empowering platform administrators to promptly ad- dress suspicious activities and mitigate potential risks. Furthermore, the system's ability to adapt pricing dynami- cally based on current market dynamics ensures competitive- ness and fairness across listings, enhancing overall user ex- perience. Through meticulous data preprocessing and feature engineering, the system effectively cleansed and transformed raw data into actionable insights, facilitating precise price predictions and anomaly detection.The implementation of this system not only strengthens the marketplace's security posture but also fosters trust and confidence among users, safeguarding their transactions and interactions. By continuously refining and optimizing its algo-rithms, the system remains resilient against emerging threats and vulnerabilities, ensuring long-term viability and efficacy. Ultimately, the project's outcome represents a significant step forward in enhancing the integrity and resilience of online marketplaces, contributing to a safer and more transparentdigital commerce ecosystem for all stakeholders involved.

VII. FUTURE SCOPE

1) E-commerce Platforms: Implementing the system on e- commerce websites can enhance fraud detection capa- bilities, ensuring the integrity of online transactions and protecting both buyers and sellers from fraudulent activ- ities.
2) Financial Services: The system could be integrated into financial platforms to monitor transactions and detect anomalies in pricing or trading behavior, aiding in the prevention of financial fraud and market manipulation.
3) Supply Chain Management: By analyzing pricing data and identifying irregularities, the system can contributeto supply chain optimization, enabling businesses to make informed decisions regarding inventory manage- ment, pricing strategies, and supplier relationships
4) Online Advertising: Incorporating the system into digital advertising platforms can help identify discrepancies in ad pricing and performance metrics, enhancing trans- parency and accountability in online advertising cam- paigns.
5) Retail Industry: Retailers can leverage the system to optimize pricing strategies, detect fraudulent activities such as price gouging or counterfeit listings, and improve overall competitiveness in the market.
6) Healthcare Sector: In healthcare, the system can assist in monitoring medical billing and pricing practices, identi- fying discrepancies that may indicate fraudulent billing or insurance fraud.
7) Travel and Hospitality: By analyzing pricing trends and detecting anomalies in hotel and airline bookings, thesystem can help prevent price manipulation and ensure fair pricing for travelers

In its whole effect, the dynamic pricing methods in online marketplace can prove as an asset for both the vendors and buyers. In future scope, it can be occupied with more accurate prediction models and utilised.

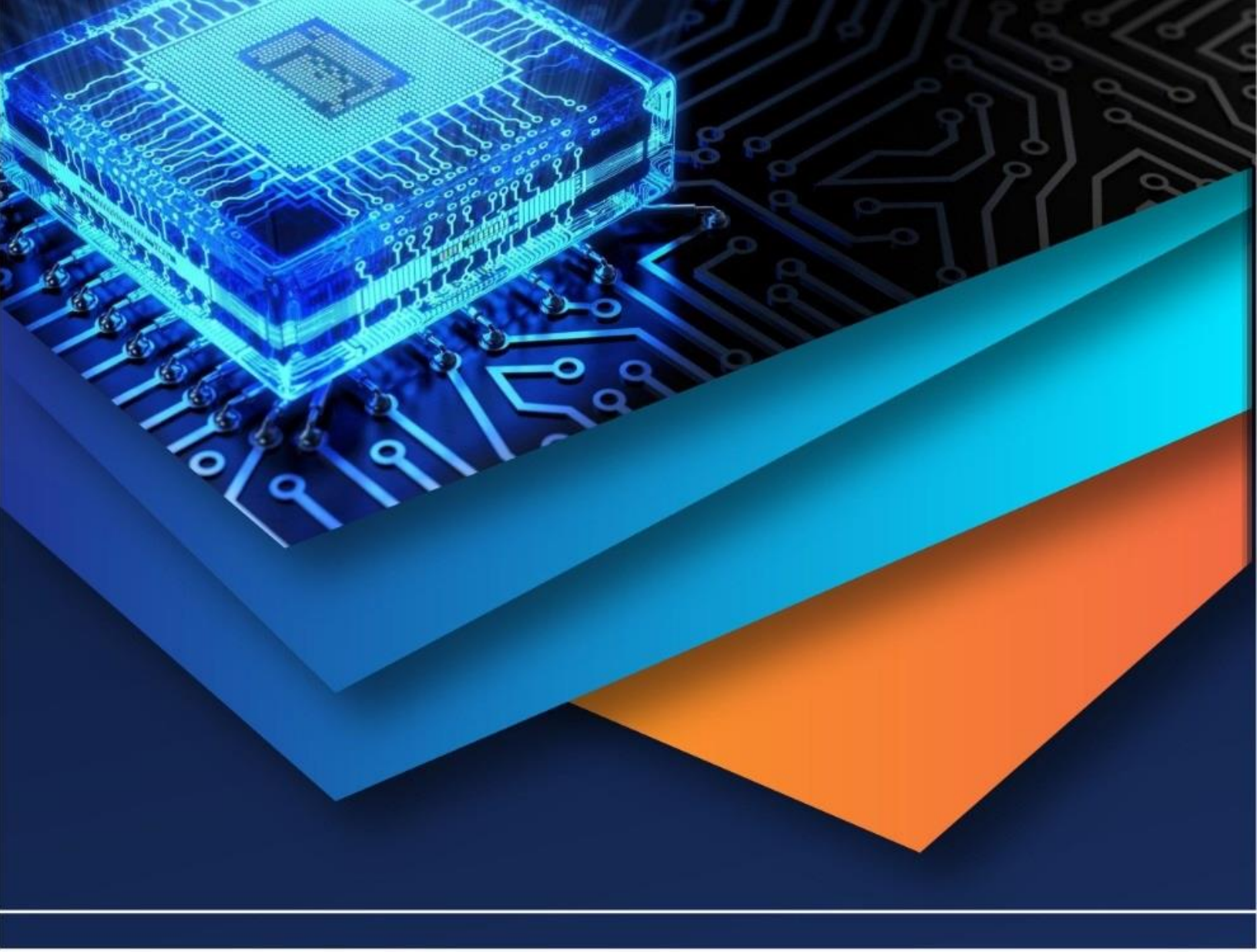
VIII. CONCLUSION

In conclusion, the dynamic pricign system for price predic- tion and fraus detection aims to mitigate the unreliable pricingstrategies used in price listing of e-commerce domains.It could potentially create a trustworthy environment and foster vendor validation feedback among consumers. In the dynamic pricingsystem, the vendors customize their prices according to the algorithm and the buyers can check their transparency. The present dynamic pricing algorithm in use is a profit oriented one whether, not the vendors but the e-commerce platforms customise their product price listings according to each user, this method aims at extracting different maximum profits according to each user. This project is essentially a reverseof the profit oriented dynamic pricing by putting forward a method for transparency oriented dynamic pricing.

REFERENCES

[1] Y. Zhang, F.Y.L. Chin, H.F. Ting Competitive algorithms for online pricing B. Fu, D.Z. Du (Eds.), Computing and Combinatorics, Lecture Notes in Computer Science, vol. 6842, Springer, Berlin, Heidelberg(2011), pp. 391-401
[2] Dynamic pricing: Definition, implications for managers, and future research directions PK Kopalle, K Pauwels, LY Akella, M Gangwar Journal of Retailing, 2023 - Elsevier
[3] Fraud detection and prevention in e-commerce: A systematic literature re- view VF Rodrigues, LM Policarpo, DE da Silveira. . . - Electronic Commerce . . . , 2022 - Elsevier
[4] Dynamic Decision-Making in Fresh Products Supply Chain With Strate-gic Consumers Zhao, Zhong, and Xinglei Chi. "Dynamic Decision-Making in Fresh Products Supply Chain with Strategic Consumers." IEEE Access (2023).
[5] How does heterogeneous consumer behavior affect pricing strategies of retail- ers? Li, Hao, and Ting Peng. "How does heterogeneous consumerbehavior affect pricing strategies of retailers?." IEEE Access 8 (2020): 165018-165033.
[6] Leveraging Product Characteristics for Online Collusive Detection in BigData Transactions Luo, Suyuan, and Shaohua Wan. "Leveraging productcharacteristics for online collusive detection in big data transactions." IEEE Access 7 (2019): 40154-40164.

[7] The role of big data and predictive analytics in retailing ET Bradlow, M Gangwar, P Kopalle, S Voleti - Journal of retailing, 2017 - Elsevier
[8] Value creation in an algorithmic world: Towards an ethics of dynamic pricing D Nunan, ML Di Domenico Journal of Business Research, 2022•Elsevier 44
[9] The impact of external reference price on consumer price expectations Au- thor links open overlay panelPraveen K Kopalle a 1, Joan Lindsey- Mullikin
[10] The Role of Big Data and Predictive Analytics in Retailing Author links open overlay panelEric T. Bradlow a, Manish Gangwar b, Praveen Kopalle c 1, Sudhir Voleti
[11] Dynamic pricing and learning: Historical origins, current research, and new directions



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY