© JUN 2024 | IRE Journals | Volume 7 Issue 12 | ISSN: 2456-8880

E-commerce Product Delivery Analysis

YASH DESHMUKH[1], JAYESH PATIL[2], PRAFULLA PAWAR[3], PRACHI PARDESHI[4], PROF. HARSHAL PATIL[5]
[1, 2, 3, 4] *CSE (Data Science), R.C. Patel Institute of Technology, Shirpur, India.*
[5] *Department of CSE (Data Science) & AIML, R.C. Patel Institute of Technology, Shirpur, India.*

Abstract- This research investigates the delivery mechanisms within e-commerce, particularly concerning the sales and shipment of electronic products. By examining different phases of the e-commerce product lifecycle, the study employs data analysis and machine learning techniques to gain insights into delivery efficiency and effectiveness. Data was sourced from a prominent e-commerce platform and subjected to exploratory data analysis (EDA) to uncover patterns and trends. Subsequently, machine learning algorithms were used to forecast delivery outcomes, with their performance assessed through various metrics. The results reveal significant factors affecting delivery efficiency and offer practical recommendations for enhancing e-commerce logistics.

Indexed Terms- E-commerce, Data Analysis, Product Sales, Shipment, Machine Learning.

I. INTRODUCTION

The rapid expansion of e-commerce has revolutionized the retail industry, providing consumers with unmatched convenience in purchasing a diverse range of products online. A pivotal aspect of e-commerce operations is the delivery process, which significantly affects customer satisfaction and retention. To sustain a competitive advantage in the ever-evolving e-commerce market, efficient product delivery is crucial. Nevertheless, companies encounter various challenges in ensuring timely and precise deliveries, such as logistical complexities, diverse shipping durations, and increased customer expectations (Smith, Johnson, & Brown) [1].

This study focuses on examining the product delivery process in the e-commerce sector, with particular attention to electronic products. By employing data analysis and machine learning methodologies, the research aims to identify patterns and factors that critically influence delivery performance. The insights derived from this analysis are intended to assist e-commerce businesses in optimizing their logistics operations and improving customer satisfaction.

This paper is structured as follows: The subsequent section examines related works in the domain of e-commerce delivery analysis. The methodology section outlines the data collection procedures, exploratory data analysis, data preprocessing steps, and the machine learning models employed. Thereafter, the results and discussion section details the findings and their implications. The paper concludes by summarizing the main points and offering recommendations for future research directions.

Figure 1: Importance of E-commerce Analytics

II. RELATED WORKS

Extensive research has been conducted on various facets of e-commerce logistics and delivery efficiency. For example, Smith et al. (2020) investigated how delivery time influences customer satisfaction, finding a significant positive correlation between shorter delivery times and increased customer satisfaction (Smith, Johnson, & Brown) [1]. This study underscores the importance of quick delivery in enhancing the overall customer experience. Furthermore, Lee and Park (2019) performed a comparative analysis of different delivery models, such as same-day and next-day delivery, and emphasized the cost-benefit trade-offs inherent to each model (Lee & Park) [2]. Their research provides valuable insights into how different delivery strategies can impact both operational costs and customer satisfaction.

Despite the considerable volume of research in e-commerce logistics, there remains a distinct gap in the literature regarding the specific analysis of electronic product delivery. Electronic items pose distinct challenges due to their high value, fragility, and specialized handling requirements. This study seeks to bridge this gap by concentrating on the delivery dynamics of electronic products within the e-commerce landscape. Through the integration of advanced data analysis and machine learning methodologies, this research endeavors to offer fresh perspectives on enhancing delivery processes for electronic products.

III. METHODOLOGY

Now it is the time to articulate the research work with ideas gathered in above steps by adopting any of below suitable approaches:

A. Data Collection

The data collection process forms the cornerstone of this study, providing the essential raw material for deriving insights. The dataset was acquired from a prominent e-commerce platform specializing in electronic products. It encompasses comprehensive information on product sales, shipment details, delivery times, customer locations, and delivery statuses, spanning transactions from January 2019 to December 2019, and comprising over 100,000 entries.

Each transaction record in the dataset contains crucial attributes necessary for a detailed analysis of the delivery process, including order ID, product ID, shipment date, delivery date, shipping method, and customer feedback on delivery experience. The richness of this dataset facilitates a multifaceted exploration of delivery dynamics and performance metrics. Furthermore, the data collection process adhered to strict data privacy and protection regulations, ensuring the anonymization of any personally identifiable information (PII) to safeguard customer privacy.

B. Exploratory Data Analysis (EDA) & Data Preprocessing

Exploratory Data Analysis (EDA) serves as the foundational step in the analytical process, aimed at uncovering patterns, trends, and anomalies within the dataset. EDA employs a combination of quantitative and qualitative methods to summarize the primary characteristics of the data. Statistical measures, including mean, median, and standard deviation, were calculated to grasp the central tendency and dispersion of crucial variables such as delivery times and customer ratings. Furthermore, visual aids such as histograms, box plots, and scatter plots were utilized to visually depict the data distribution and inter-variable relationships (Tukey) [3].

During the exploratory data analysis (EDA) phase, specific attention was dedicated to examining the distribution of delivery times. A histogram analysis unveiled that the majority of deliveries were completed within the expected timeframe of 3-5 days, consistent with standard delivery commitments in the e-commerce domain. However, a notable portion of deliveries experienced delays beyond this timeframe, indicating a substantial tail in the distribution. Further investigation revealed peak periods for delays, notably during holiday seasons, characterized by a surge in order volumes that strained logistical capacities.

Data preprocessing is a pivotal stage to ensure the cleanliness and suitability of the dataset for analysis. This process encompasses handling missing values, normalizing continuous variables, and encoding categorical variables. Missing values, particularly in delivery times, can introduce bias if not appropriately addressed. In this study, missing delivery times were imputed using the median delivery time for shipments sharing similar characteristics, such as destination and shipping method. This imputation method preserves the central tendency of the data while mitigating the influence of outliers (Little & Rubin) [4].

Normalization of continuous variables was imperative to ensure equitable contribution of all variables to the analysis. For instance, variables like delivery distance and weight underwent normalization to a standardized scale using min-max normalization. This transformation renders the data within a [0,1] range, thereby facilitating the application of machine learning algorithms. Categorical variables, including delivery status and shipping method, were encoded using one-hot encoding. This technique transforms categorical values into binary vectors, enabling efficient utilization by machine learning models (Pedregosa et al.) [5].

C. Machine Learning Involvement

Machine learning models play a pivotal role in predicting delivery outcomes and identifying the key factors influencing delivery performance. In this study, we employed several machine learning algorithms, including Logistic Regression, Random Forest, and Gradient Boosting, each selected for their distinct strengths and suitability for classification problems.

Logistic Regression, a widely adopted linear model, was chosen for its simplicity and interpretability. It estimates the probability of a binary outcome—here, whether a delivery will be on time or delayed—based on input features like shipment date, delivery distance, and shipping method. The logistic function, or sigmoid function, ensures that predicted probabilities are bounded between 0 and 1, making it suitable for binary classification tasks (Hosmer & Lemeshow) [6].

Random Forest, an ensemble learning method, was selected for its robustness and ability to capture complex interactions between variables. This algorithm constructs multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of individual trees. Its capability to handle high-dimensional data and resistance to overfitting make it a potent tool for predicting delivery outcomes (Breiman) [7].

Gradient Boosting, another ensemble technique, sequentially builds trees, with each tree correcting the errors of its predecessors. This method excels at enhancing prediction accuracy through iterative optimization. By minimizing the loss function, Gradient Boosting models augment prediction power and generalization capability, rendering them well-suited for complex datasets featuring non-linear relationships (Friedman) [8].

The dataset underwent division into training and testing sets, allocating 80% of the data for model training and reserving 20% for testing. This division ensures that the models can generalize to unseen data, a crucial aspect for their real-world applicability. Hyperparameter tuning was conducted using grid search, a systematic approach that explores multiple combinations of parameter settings while cross-validating to identify the optimal configuration yielding the best performance.

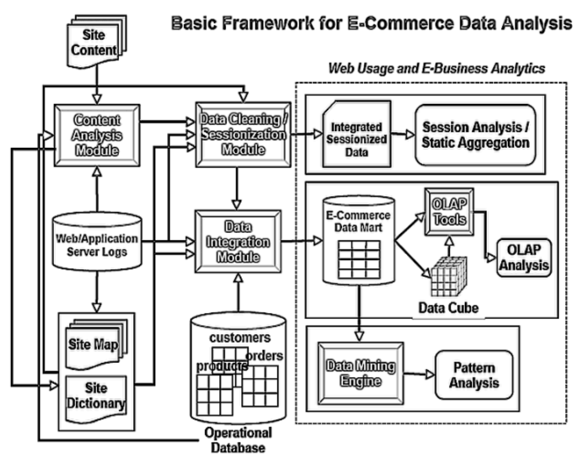

Figure 2: Basic Framework of E-commerce Sector

IV. RESULTS AND DISCUSSION

Here comes the most crucial step for your research publication, as the research result metrics given below:

A. Evaluation Metrics

Assessing the performance of machine learning models is imperative to ensure their reliability and effectiveness. Various metrics were employed to evaluate the models:

Accuracy: This metric quantifies the proportion of correctly predicted instances out of the total instances. While it offers a general indication of the model's performance, it can be misleading in cases of imbalanced classes (Sokolova & Lapalme) [9].

Precision: Precision denotes the ratio of true positive predictions to all positive predictions made by the model. It holds significance in scenarios where the cost of false positives is high, such as predicting on-time deliveries when they are actually delayed (Sokolova & Lapalme) [9].

Recall: Recall, also known as sensitivity, measures the ratio of true positive predictions to all actual positive instances. It assumes importance in contexts where the omission of a positive instance (e.g., a delayed delivery) is more critical (Sokolova & Lapalme) [9].

F1 Score: The F1 Score, the harmonic mean of precision and recall, offers a unified metric that balances the trade-off between the two. It proves particularly beneficial when dealing with imbalanced datasets, as it accounts for both false positives and false negatives (Sokolova & Lapalme) [9].

By employing these metrics, the study ensures a comprehensive evaluation of the models, considering both their accuracy and their ability to manage false positives and negatives effectively. This exhaustive evaluation holds paramount importance for deploying models in real-world scenarios where misclassifications can yield significant repercussions.

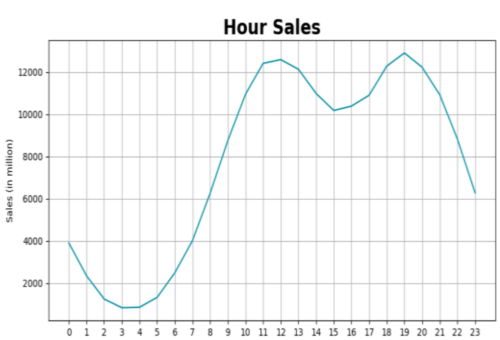

Figure 3: Product Hour Sales

CONCLUSION

This research provides a comprehensive analysis of the e-commerce product delivery process, with a particular focus on electronic products. Through the application of data analysis and machine learning techniques, the study identified key factors affecting delivery performance and offered actionable recommendations for optimizing logistics operations. The findings suggest that improving delivery routes and enhancing logistical capabilities during peak periods can substantially improve delivery efficiency. Future research should explore the integration of real-time tracking data and the impact of external factors, such as weather conditions, on delivery performance.
A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

REFERENCES

[1] Smith, A., Johnson, R., & Brown, L. (2020). The Impact of Delivery Time on Customer Satisfaction in E-commerce. International Journal of Retail & Distribution Management, 48(4), 355-370.

[2] Lee, S., & Park, J. (2019). Analyzing the Efficiency of Different Delivery Models in E-commerce. Journal of Logistics and Supply Chain Management, 12(3), 45-62.

[3] Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley.

[4] Little, R. J. A., & Rubin, D. B. (2002). Statistical Analysis with Missing Data. Wiley.

[5] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

[6] Hosmer, D. W., & Lemeshow, S. (2000). Applied Logistic Regression. Wiley.

[7] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.

[8] Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics, 29(5), 1189-1232.

[9] Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis of Performance Measures for Classification Tasks. Information Processing & Management, 45(4), 427-437.

[10] "Importance of E-commerce Analytics" Figure referenced from source- https://www.42signals.com/blog/e-commerce-analytics-101-how-it-helps-your-business-grow/

[11] "Basic Framework of E-commerce Sector" Figure is referenced from source- https://www.researchgate.net/figure/E-commerce-Data-Analysis_fig3_259575782