International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
Article Received: *25 December 2022* ***Revised:*** *12 January 2023* ***Accepted:*** *20 February 2023*

Leveraging Machine Learning for Predictive Analytics in Ecommerce

Ramkumar Soundarapandian.

Senior Manager - Capgemini America Inc, 333 W Wacker Dr #300, Chicago, IL 60606, United States of America.

ramkumarcg@gmail.com

ABSTARCT

Predictive analytics is becoming more and more necessary for organizations to use in the quickly changing e-commerce industry in order to predict customer behaviour, optimize marketing campaigns, and improve overall operational efficiency. The goal of this research study is to strengthen predictive analytics in the e-commerce industry by utilizing machine learning approaches.AI is essential for the business analyst's ability to make predictions. AI is a rapidly developing field that is employed in all fields, particularly in data analysis and prediction. In this study, machine learning—a subset of AI—is used for this purpose. The paper's goal is to determine the potential for growth and application of e-commerce in the future. In order to provide a clear explanation, we have used secondary data on the market value of e-commerce as a basis to support the income generated by e-commerce, and online consumers are taken to understand the total contribution of the e-commerce sector in India, then attempted to use Python to determine the forecast for the following years, 2024 to 2026. The current study aims to improve the accuracy of market value projection by using two more factors: the percentage of Indian online shoppers and e-commerce revenue. This paper attempts to offer practical suggestions and best practices for e-commerce practitioners and decision-makers wishing to leverage machine learning for predictive analytics by combining insights from both academic research and industry operations. In the end, the study advances our understanding of e-commerce analytics and establishes the groundwork for more in-depth investigation and creative thinking in this area.

Keywords: Machine Learning, Predictive analysis, Ecommerce, Customer behaviour, Data quality, Fault Detection.

1. INTRODUCTION

Businesses always look for novel approaches to obtain a competitive edge, improve consumer experiences, and maximize operational efficiency in the quickly changing field of e-commerce (Patel, 2020). Predictive analytics using machine learning (ML) has become a potent tool in this quest, transforming supply chain operations, forecasting demand, personalizing suggestions, and helping firms predict customer behaviour (Adrian Micu, 2021). This study explores the uses, advantages, difficulties, and potential consequences of using machine learning techniques to predictive analytics in e-commerce.

Massive volumes of data have been produced by the exponential growth of e-commerce platforms, ranging from competitor studies and industry trends to browsing and purchase history of individual customers (Morsi, 2020). Such vast and complicated data sets are typically difficult for traditional analytics techniques to glean useful insights from (Kotsokechagia, 2021). But by enabling automated data processing, pattern identification, and predictive modelling, machine learning algorithms provide an answer (Nanduri, 2020). Machine learning algorithms have the ability to analyze historical data to find patterns, make accurate predictions about the future, and find correlations (Narayana, 2021).

Customer behaviour analysis and segmentation is one of the main uses of machine learning in e-commerce (Rajesh, 2021). Businesses can learn more about the preferences, spending patterns, and lifecycle stages of their customers by utilizing machine learning algorithms (Kharfan, 2021). This makes it possible to create highly focused marketing campaigns, customized promotional offers, and personalized product recommendations, all of which improve consumer engagement and increase sales (Pawłowski, 2022). Additionally, proactive client retention techniques are made possible by machine learning-powered predictive analytics, which may identify at-risk consumers and take preventative action to reduce churn (Saleem, 2019).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
Article Received: *25 December 2022* ***Revised:*** *12 January 2023* ***Accepted:*** *20 February 2023*

1.1. The Rise of Predictive Analytics in E-Commerce: An Overview

Predictive analytics' birth and quick rise in the e-commerce industry reflect a dramatic shift in how companies use data to understand and forecast customer behaviour, preferences, and market trends (Sharma, 2022). Historically, decision-making in e-commerce operations was based on historical data and crude forecasting techniques. But now that sophisticated machine learning algorithms and big data technologies have been developed, companies are able to use massive volumes of both structured and unstructured data to quickly derive actionable insights (Matuszelański, 2022). In e-commerce, predictive analytics involves using complex algorithms to examine past transaction histories, user interactions, browsing patterns, demographic information, and other relevant variables in order to predict future trends, identify possible hazards, personalize marketing strategies, enhance inventory control, and improve overall customer experiences (Zhang, 2021). E-commerce companies can use predictive analytics to forecast demand properly, eliminate inventory-related risks, customize product recommendations, optimize pricing tactics, and anticipate client needs (Wang, 2021).

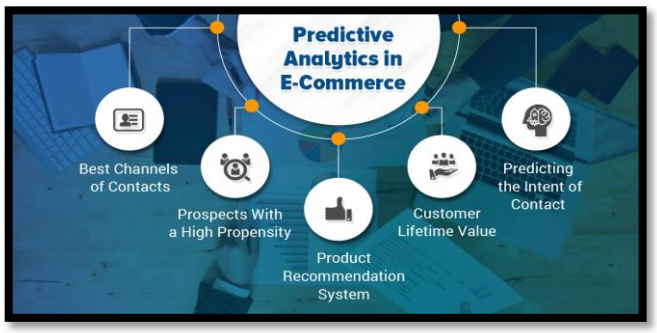

Figure 1: E-Commerce Predictive Analytics

1.2. Objectives of the Study

The following is a summary of this study's main goals:

- To project the Indian e-commerce industry's market worth.
- To increase understanding of the importance and future prospects of India's e-commerce market.

2. LITERATURE REVIEW

Bawack et al. (2022) we out a thorough literature research and bibliometric analysis on the application of artificial intelligence (AI) in e-commerce. The research explores the rapidly changing field of AI applications in e-commerce, emphasizing significant developments, obstacles, and prospects. The authors offer insights on the application of AI techniques including machine learning, natural language processing, and recommendation systems in improving several facets of E-commerce operations by examining a wide range of academic articles (Bawack, 2022).

Iqbal (2022) examines the various uses of machine learning in the context of online shopping. The study explores the numerous ways that machine learning techniques are being used to improve various elements of e-commerce operations via the lenses of Organization, Business, and Management. Iqbal offers insightful information about how machine learning applications are changing business strategy, enhancing customer experiences, and enhancing decision-making in the e-commerce industry by analyzing the state of the field (Iqbal, 2022).

Micu et al. (2019) by investigating the application of machine learning techniques to improve e-commerce performance, you can add to the body of knowledge. Their research on the use of machine learning approaches to enhance many aspects of e-commerce operations was published in the Annals of Dunarea Jos University of Galati. The authors hope to increase the overall efficacy and efficiency of e-commerce endeavours by optimizing e-commerce performance measures like fraud detection, recommendation systems, consumer segmentation, and sales forecasting through the use of machine learning algorithms (Micu, 2019).

Policarpo et al. (2021) carry out an extensive systematic assessment of the literature to offer a current knowledge of machine learning's function in e-commerce. Their research provides insightful information on distinct machine learning technologies and how they affect various facets of e-commerce operations. Policarpo and colleagues demonstrate the wide spectrum of machine learning (ML) approaches used in e-commerce, such as recommendation systems, predictive analytics, fraud detection, and personalized marketing, by carefully examining the body of existing research. They also highlight the vital role that data plays in powering machine learning algorithms, stressing the significance of data security, privacy, and quality in e-commerce environments. Policarpo et al. provide practitioners and researchers with a thorough overview of the state of machine learning (ML) in e-commerce by combining the results of multiple studies and highlighting new trends, problems, and areas for future research (Policarpo, 2021).

Yeung et al. (2019) investigate how to use machine learning (ML) in data analytics for cloud-based e-commerce. The authors examine the ways in which machine learning and cloud computing might work together to improve the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
Article Received: *25 December 2022* ***Revised:*** *12 January 2023* ***Accepted:*** *20 February 2023*

efficiency, agility, and scalability of e-commerce operations. Yeung and colleagues illustrate the usefulness of combining machine learning (ML) algorithms with cloud-based data analytics tools using a case study method. This allows for real-time decision-making, dynamic pricing optimization, and customized consumer experiences. They also talk about how ML-driven e-commerce systems can be made even more powerful by utilizing distributed computing architectures like serverless and edge computing. Yeung et al. offer firms looking to leverage these technologies to obtain a competitive advantage in the digital marketplace insightful explanations of the technical nuances and pragmatic issues of ML deployment in cloud-based e-commerce systems (Yeung, 2019).

3. RESEARCH METHODOLOGY

This section explains the study's methodology, which included data gathering, analysis methods, and the use of artificial intelligence (AI) to make predictions.

3.1. Data Collection

The study mostly makes use of secondary data that was gathered from various sources, with a focus on using statista.com data. Secondary data provide a multitude of information relevant to the research goals, allowing for a thorough examination of trends and patterns in the e-commerce industry.

3.2. AI Algorithm: Linear Regression

One basic AI method used to determine correlations between two variables, one dependent and the other independent, is called linear regression. In order to make predictions about the dependent variable based on changes in the independent variable, the goal is to fit a linear equation to the data. The following is the formula for a basic linear regression model:

$$y = b0 + b1 * x$$

'y' stands for the dependent variable, 'x' for the independent variable, 'b0' for the line's intercept, and 'b1' for the line's slope in this equation. When combined, these elements allow the dependent variable to be predicted for fresh values of the independent variable.

3.3. Research Architecture

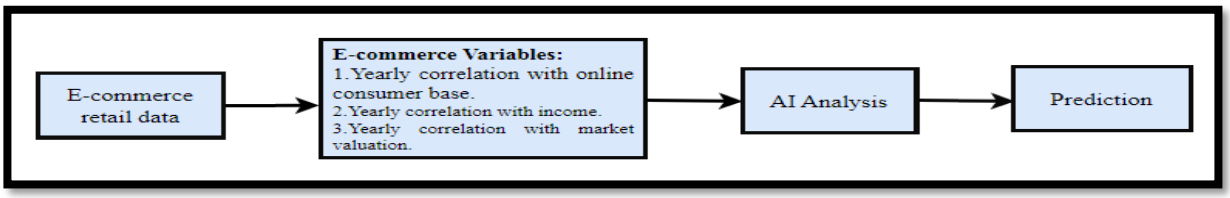

Figure 2: Research Architecture

The process flow of the study is outlined by the research architecture, which is depicted in Figure 2. The study specifically focuses on gaining insights from retail e-commerce data, highlighting important variables like Online Shoppers, Revenue, and Market Value. These variables provide an organized framework for examining trends and dynamics in the e-commerce industry by acting as focal points for AI analysis and predictive modelling.

3.4. Data Analysis

Python analytics tools, which provide a stable foundation for data processing, visualization, and modelling, are used to examine the gathered data. Because of its adaptability and wide library support, Python is a great option for doing complex data analysis and extracting valuable insights from large, complicated datasets..

3.5. Integration of AI for Predictive Analysis

Predictive modelling and forecasting are made possible with the integration of artificial intelligence (AI) algorithms into the analysis process. In particular, the study makes use of the statistical modelling approach known as linear regression to determine correlations between important factors in the e-commerce industry.

4. DATA ANALYSIS AND RESULTS

This study uses the value of the e-commerce market as its base, with the addition of two other parameters for support. Following data collection, analysis is done to forecast future years for India's e-commerce sector. To evaluate and forecast the third element, the first two factors listed below are used.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
Article Received: *25 December 2022* ***Revised:*** *12 January 2023* ***Accepted:*** *20 February 2023*

1. Yearly correlation with online consumer base

To forecast the future of India's e-commerce sector and the number of online shoppers in the nation, data covering the years 2019 to 2022 has been collected from Statista's website. This report can assist retailers become more aware of the power and popularity of online buying in India. Merchants can better comprehend the potential of online shoppers and modify their strategy by anticipating and making recommendations based on this data.

The steps utilized for the study and forecasting of Indian online consumers are as follows.

The year (which might be any year between 2019 and 2022) and the percentage of online shoppers are the two variables in the dataset. I got the data from the Statista website, and I'll use predictive analysis with it in my research. Several sample data points from the dataset are shown here.

Table 1: Year and Online Shopper Sample in the Millions

Year	Online Shoppers
2019	7.10
2020	7.50
2021	8.20

Following the import of all the data, the data are displayed using the year and the number of online shoppers as the x and y axes, respectively, in a straightforward graphical style.

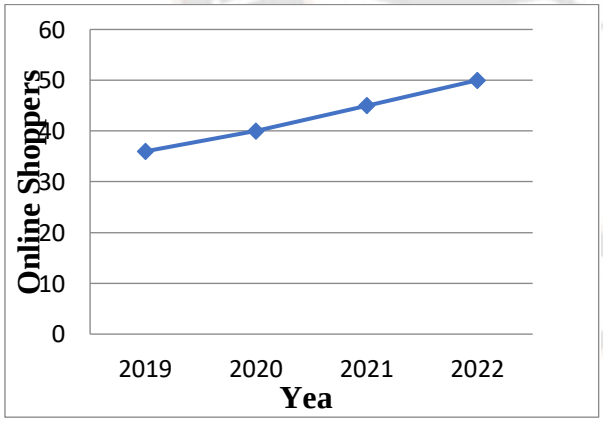

Figure 3: Year and Online Shoppers

Following the import of all the data, the data are displayed using the year and the number of online shoppers as the x and y axes, respectively, in a straightforward graphical style.

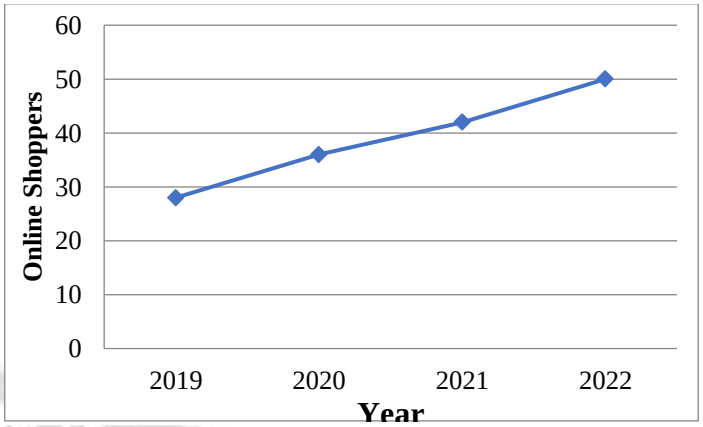

Figure 4: Accuracy – Year and Online Shoppers

Figure 3 above illustrates the 94 percent accuracy rate for the prediction based on the factor of online shoppers. The accuracy rate has been determined by plotting the same.

Following its AI training, the model will determine the prediction for the upcoming years.

```
onlineshoppers_2024 = model.predict(year_2024)
onlineshoppers_2025 = model.predict(year_2025)
onlineshoppers_2026 = model.predict(year_2026)
print(" Prediction made on Online Shoppers in Millions")
print("2024:", onlineshoppers_2024[0])
print("2025:", onlineshoppers_2025[0])
print("2026:", onlineshoppers_2026[0])
Prediction made on Online Shoppers in Millions
2024: 67.47239532
2025: 63.83518672
2026: 67.47239532
```

Image 1: Screenshot of the Program Prediction: Year and Online Shoppers

Here, it is anticipated that the number of online shoppers would rise to 65 million in 2024, slightly decline to 62 million in 2025, and then rise to 65 million once more in 2026. Given its high accuracy rate of 94%, the prediction is considered dependable.

2. Yearly correlation with income

A few years and e-commerce income have been taken into account for analysis and forecasting in order to determine the next factor. Some of the sample data sets that are imported for analysis are shown below.

Table 2: An example of year and million-dollar e-commerce revenue statistics

Year	Ecommerce Revenue
2019	22.03
2020	26.96

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
Article Received: *25 December 2022* ***Revised:*** *12 January 2023* ***Accepted:*** *20 February 2023*

2021	33.21
2022	40.11

After the model has been trained using the tested data set, it is claimed that the data can be used for prediction. A matching algorithm can be used to forecast how many will occur each year. Here, the x and y axes will represent the year and e-commerce revenue, respectively, for the test data. The plotting example that is displayed is as follows:

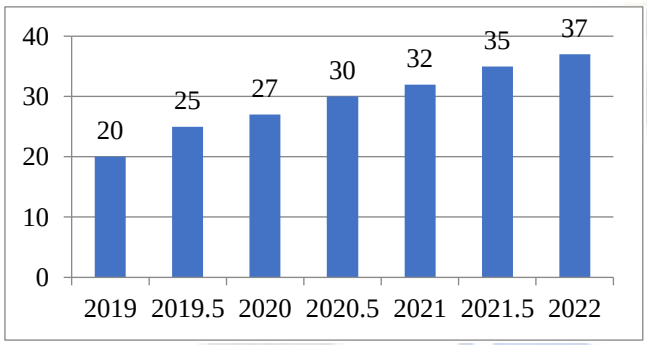

Figure 5: Accuracy – Year and Ecommerce Revenue

The model made predictions for the years 2024, 2025, and 2026, and when compared to the prior factor, its accuracy rate (98%) was also good.

```
revenue_2024 = model.predict(year_2024)
revenue_2025 = model.predict(year_2025)
revenue_2026 = model.predict(year_2026)
print(" Prediction made on Ecommerce Revenue in Millions")
print("2024:", revenue_2024[0])
print("2025:", revenue_2025[0])
print("2026:", revenue_2026[0])


 Prediction made on Ecommerce Revenue in Millions
2024: 59.83999999999833
2025: 65.34499999999935
2026: 70.84999999999854
```

Image 2: Screenshot of the Program Prediction: Year and E-Commerce Revenue

According to reports, there will be 60 million people in 2024, 66 million in 2025, and 71 million in 2026 has a respectable 99% accuracy rate. Therefore, e-commerce income in India is expected to grow, which encourages merchandisers to go online and reach clients at their leisure.

3. Yearly correlation with market valuation

To evaluate the market value for the current year and the years to come in the Indian e-commerce sector. These few data were used to arrive at the conclusions.

Table 3: Sample data by year and market value in millions

Year	Market Value
2019	22.02
2020	25.02
2021	27.02
2022	29.02

Plotting is done utilizing all of the training data after the model has been tested and trained on the test data set. This helps determine how accurate the projected data is.

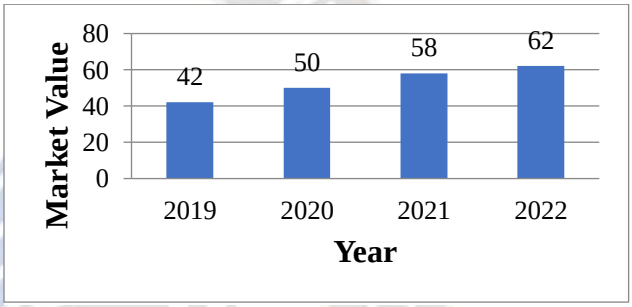

Figure 6: Accuracy – Year and Market Value

The model forecasted the years 2024, 2025, and 2026, and when compared to the prior factor, its accuracy rate (91%) was likewise good.

```
marketvalue_2024 = model.predict(year_2024)
marketvalue_2025 = model.predict(year_2025)
marketvalue_2026 = model.predict(year_2026)
print(" Prediction made on Ecommerce Market Value in Millions")
print("2024:", marketvalue_2024[0])
print("2025:", marketvalue_2025[0])
print("2026:", marketvalue_2026[0])
Prediction made on Online Shoppers in Millions
2024: 67.67391304
2025: 73.56956522
2026: 79.4621739
```

Image 3: Screenshot of the Program Prediction: Year and Market Value

And the last element According to market value predictions for e-commerce in India, the percentage will continue to rise through 2026. With a 92% accuracy rate, it is 68 million for 2024, 74 million for 2025, and 80 million for 2026. Therefore, it makes sense for the merchandiser to transition to an online store.

5. CONCLUSION

Predictive analytics using machine learning in e-commerce is a major development in contemporary commercial methods. Businesses can now obtain priceless insights into consumer behaviour, market trends, and operational efficiencies by utilizing sophisticated algorithms and large datasets. This study solely looked at e-commerce, which is considered to be an emerging market. We added two more elements to support the suggestions on the future scope of ecommerce in the merchandising industry in order to determine the specific market worth in India. Online shopping benefits both buyers and sellers by making it easier for them to connect with potential clients. This study accurately estimated the size and significance of the e-commerce market in the future, demonstrated that the industry gives more to society than it does to the economy through revenue projections, and estimated the density of future online customers, which raises awareness among sellers. As a result, it can be argued that the study accurately captures the scope and significance of the e-commerce business through the analysis and prediction of artificial intelligence. It also suggests that the merchandising sector shift to online retailing. The key thing for businesspeople to remember is to apply all kinds of innovations according to their business type and to remember that every firm should ultimately end up contributing more to India's revenue, even though the prognosis and conclusion have been reached.

REFERENCES

1. Adrian Micu, Angela-Eliza Micu, et.al (2021) issued in Artificial Intelligence in Wholesale and Retail, DOI: 10.24818/EA/2021/56/137
2. Bawack, R. E., Wamba, S. F., Carillo, K. D. A., & Akter, S. (2022). Artificial intelligence in E-Commerce: a bibliometric study and literature review. Electronic markets, 32(1), 297-338.
3. Iqbal, M. (2022). Machine learning applications in e-commerce. Organization, Business and Management, 65.
4. Kharfan, M., Chan, V. W. K., & Firdolas Efendigil, T. (2021). A data-driven forecasting approach for newly launched seasonal products by leveraging machine-learning approaches. Annals of Operations Research, 303(1), 159-174.
5. Kotsokechagia, M. (2021). Predictive model for customer satisfaction in e-commerce.
6. Matuszelański, K., & Kopczewska, K. (2022). Customer churn in retail e-commerce business: Spatial and machine learning approach. Journal of Theoretical and Applied Electronic Commerce Research, 17(1), 165-198.
7. Micu, A., Geru, M., Capatina, A., Avram, C., Rusu, R., & Panait, A. A. (2019). Leveraging e-Commerce performance through machine learning algorithms. Ann. Dunarea Jos Univ. Galati, 2, 162-171.
8. Morsi, S. (2020). A Predictive Analytics Model for E-commerce Sales Transactions to Support Decision Making: A Case Study. International Journal of Computer and Information Technology (2279-0764), 9(1).
9. Nanduri, J., Jia, Y., Oka, A., Beaver, J., & Liu, Y. W. (2020). Microsoft uses machine learning and optimization to reduce e-commerce fraud. INFORMS Journal on Applied Analytics, 50(1), 64-79.
10. Narayana, C. V., Likhitha, C. L., Bademiya, S., & Kusumanjali, K. (2021, August). Machine learning techniques to predict the price of used cars: predictive analytics in retail business. In 2021 second international conference on electronics and sustainable communication systems (ICESC) (pp. 1680-1687). IEEE.
11. Patel, N., & Trivedi, S. (2020). Leveraging predictive modeling, machine learning personalization, NLP customer support, and AI chatbots to increase customer loyalty. Empirical Quests for Management Essences, 3(3), 1-24.
12. Pawłowski, M. (2022). Machine learning based product classification for ecommerce. Journal of Computer Information Systems, 62(4), 730-739.
13. Policarpo, L. M., da Silveira, D. E., da Rosa Righi, R., Stoffel, R. A., da Costa, C. A., Barbosa, J. L. V., ... & Arcot, T. (2021). Machine learning through the lens of e-commerce initiatives: An up-to-date systematic literature review. Computer Science Review, 41, 100414.
14. Rajesh, G., Preethi, S. P., Priya, R. S., Rajesh, L., & Raajini, X. M. (2021). Customer Behavior Prediction for E-Commerce Sites Using Machine Learning Techniques: An Investigation. Reinventing Manufacturing and Business Processes Through Artificial Intelligence, 79-93.
15. Saleem, H., Muhammad, K. B., Nizamani, A. H., Saleem, S., & Aslam, A. M. (2019). Data science and machine learning approach to improve E-commerce sales performance on social web. International Journal of Computer Science and Network Security (IJCSNS), 12(4), 401-424.

16. Sharma, M., Sharma, V., & Kapoor, R. (2022). Study of E-Commerce and Impact of Machine Learning in E-Commerce. In Empirical Research for Futuristic E-Commerce Systems: Foundations and Applications (pp. 1-22). IGI Global.
17. Vishal Dineshkumar Soni, 2020 issed in International Journal of Trend in Scientific Research and Development (IJTSRD), Volume 4 Issue 5, August 2020 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470
18. Wang, X. S., Ryoo, J. H. J., Bendle, N., & Kopalle, P. K. (2021). The role of machine learning analytics and metrics in retailing research. Journal of Retailing, 97(4), 658-675.
19. Yeung, J., Wong, S., Tam, A., & So, J. (2019, July). Integrating machine learning technology to data analytics for e-commerce on cloud. In 2019 Third World Conference on Smart Trends in Systems Security and Sustainablity (WorldS4) (pp. 105-109). IEEE.
20. Zhang, D., Pee, L. G., & Cui, L. (2021). Artificial intelligence in E-commerce fulfillment: A case study of resource orchestration at Alibaba's Smart Warehouse. International Journal of Information Management, 57, 102304.