iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** I **Month of publication:** January 2025

DOI: https://doi.org/10.22214/ijraset.2025.66732

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 13 Issue I Jan 2025- Available at www.ijraset.com

Sales Forecast and Inventory Management Using ML

Ms. Aangi Khandhar[1], Dr. Sweety Garg[2]
[1]*Student, Department of MSc.IT, Nagindas Khandwala College, Mumbai, Maharashtra, India*
[2]*Assistant Professor, Department of Computer and Information Science, Nagindas Khandwala College, Mumbai, Maharashtra, India*

Abstract: ***Efficient sales forecasting and management of inventory are some of the most major concerns for businesses in the modern economic environment, where organizations have to optimize resources with a view to reducing operational costs and meeting customers' demands. The effort is towards developing a robust predictive model capable of analyzing the trends in sales and requirements in the area of inventory by applying Machine Learning techniques so that actionable insights can be derived based on the analyzed information for proper supply chain management.***
In fact, the very raison d'être of the premise is to analyzing historical sales data for patterns that can then be used to make accurate predictions, hence enabling business firms to anticipate changes in demand and adjust inventory levels accordingly. This paper uses state-of-the-art Machine Learning algorithms-Regression Analysis and Time Series Forecasting models-to assure precise forecast accuracy. This algorithm will be able to capture seasonality, trends, and anomalies in the salesdata so that the model adapts to the changing market conditions.
Data preprocessing and feature engineering are done to clean the data and convert it into usable information that enhances the model's performance by a significant factor. Visualization techniques are carried out to show findings intuitively and interpretively to stakeholders for effective decision-making. The solution also provided recommendations on the minimum level of inventory required to be maintained so that overstocking and stockouts are at minimal, resulting in reduced losses on account of unwanted stocking or on failure of sales.
This paper points out the transformational power of Machine Learning in supply chain optimization, showing how predictive analytics can drive data-driven decision-making and improve overall operational efficiency. By accurately forecasting sales and aligning inventory management strategies, businesses can enhance customer satisfaction and maintain a competitive advantage in dynamic markets.
Keywords: Sales Forecasting, Inventory Management, Machine Learning, Time Series Analysis, Demand Prediction, Supply Chain Optimization.

I. INTRODUCTION

Among the most important ways for modern businesses to stay profitable and meet customer needs in today's competitive environment are accurate sales forecasting and efficient inventory management. Businesses face major challenges, such as fluctuating demand, overstocking, and stockouts, which can result in financial losses and missed
opportunities. These are challenges that require advanced analytical techniques that can predict sales trends and optimize stock levels.

This paper looks at the application of Machine Learning techniques in developing a predictive model for sales forecasting and inventory management. Historic sales data analysis and various algorithms, such as regression and time series analysis, identify demand patterns that create actionable insights. These enable the adjustment of inventory levels to match respective expected demand, reducing waste and storage costs, thus enhancing operational efficiency.

Besides, much attention is drawn to data preprocessing and feature engineering that help to get more accurate results of predictions, which enable visualization to present the outcome intuitively for decision-makers. The embedding of Machine Learning into sales forecasting and inventory management allows one to show its great possibility of transforming many of these conventional business processes into dynamic, information-driven operations.

Introduction The following introduction provides the basis necessary for starting to understand how Machine Learning is capable of transforming inventory control to support adaptability in business with regards to market demands:.

©IJRASET: All Rights are Reserved

II. LITERATURE REVIEW

Michael (2021), examines sales forecasting practices in 13 large Australian manufacturing companies, comparing company forecasts with naive forecasts, which use the most recent value as the estimate. Surprisingly, company forecasts were not consistently more accurate than naive forecasts, even though significant effort and contextual data were involved. Metrics such as MAPE, SMAPE, and GMRAE were used to assess accuracy, revealing that inefficiencies, biases, and serial correlation in errors impacted forecast quality. The findings highlight the need for improved forecasting methods despite the resources invested.

Smith (2019) extends more on segmentation strategies and identifies behavioral segmentation as one of the key factors in the identification of customers' needs and preferences. The work of Smith identifies that behavioral segmentation, which classifies consumers by their interaction with products and services, presents so-called golden insights into motivations and patterns of a purchase.

Brown and Green (2020) intersects where segmentation collides with behavioral analysis. Their study provides sophisticated analytical methods in consumer behavior, such as the development of forecast purchasing patterns based on Machine Learning algorithms. They advocate supplementing behavioral data with segmentation-focused strategies to stack up the effectiveness of marketing.

Miller (2018) emphasis is on the ways in which consumer behavior theories shape segmentation practices. Miller's contribution was to underline a dynamic role of the segmentation itself, where continuous monitoring and analysis of changes in behavioral aspects and consumer preferences is made.

Johnson (2022) explores how digital technologies have impacted segmentation and behavioral analysis. According to Johnson, the traditional methods of segmentation are being revitalized by digital tools and platforms in order to allow a more precise and real-time analysis of consumer behavior. This further leads to more personalized marketing strategies with improved customer engagement.

Choi and Lee (2017) reviews the role of cultural and demographic elements in segmentation. According to their study, under such integration into segmentation models, one could gain more insight into consumer behaviour and therefore targeting activity. They recommend a multi-dimensional approach to segmentation that would consider various externally-induced influences.

III. RESEARCH OBJECTIVES

1) To predict future sales using AI-based models for better inventory and strategy planning.
2) To identify key sales drivers by analyzing historical data and trends.
3) To automate sales forecasting for improved accuracy and decision-making

IV. RESEARCH METHODOLOGY

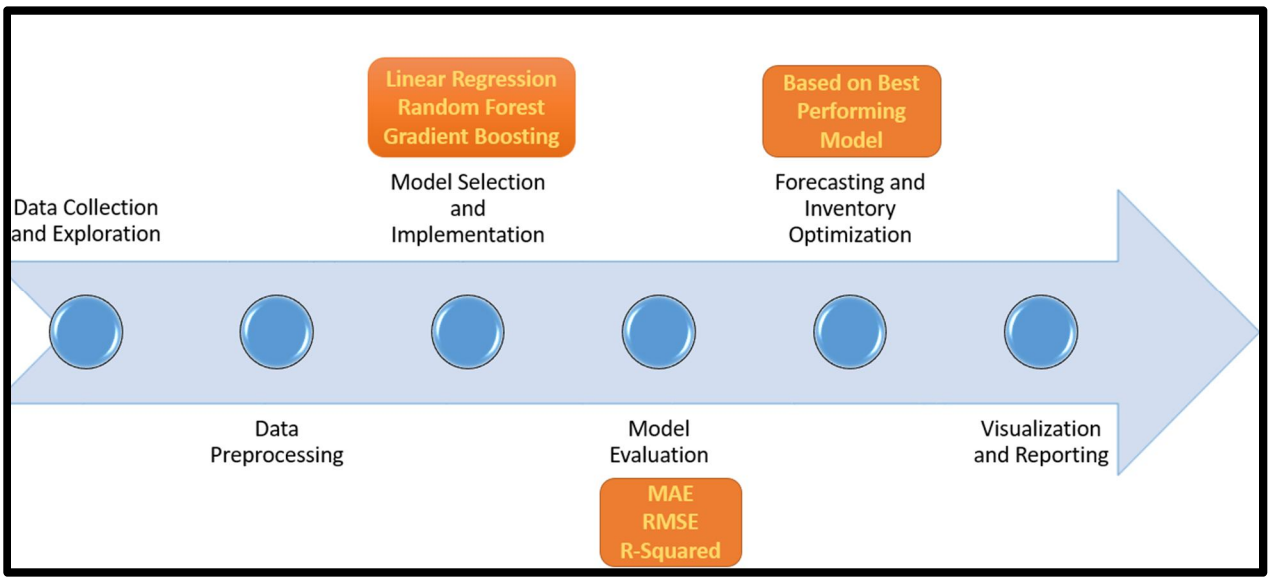


Fig 1: Methodology

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

It provides a systematic approach to analyzing and foreseeing sales by optimizing the inventory using Machine Learning techniques. It follows a proper flow of importing data, cleaning up the datasets for missing values and outliers, and any inconsistency in the data. EDA is then performed using visualization from Matplotlib and Seaborn to find patterns, trends, and correlations in the data. This prepares the data for modeling by feature scaling and selection. Applied various regression techniques on the dataset like Linear Regression, Random Forest, and Gradient Boosting, their performance is compared using MSE and R-squared metrics. Applied hyperparameter tuning using Grid Search and Randomized Search. The best model has to be deployed in predicting future sales trends to allow the making of necessary data-driven decisions for inventory. Besides, visualization of decision trees can be done to interpret the results correctly, knowing that the insights achieved are actionable toward the goal of inventory optimization. Warnings and potential errors are managed using robust Python libraries to ensure the reliability of code.

A. Sample Data

	Unnamed: 0	TV	Radio	Newspaper	Sales
0	1	230.1	37.8	69.2	22.1
1	2	44.5	39.3	45.1	10.4
2	3	17.2	45.9	69.3	9.3
3	4	151.5	41.3	58.5	18.5
4	5	180.8	10.8	58.4	12.9

Table 1: Dataset

B. Libraries used

1) NumPy: Handles numerical computations and supports array operations for efficient data processing.
2) Pandas: Provides tools for data manipulation and analysis, such as working with tables (DataFrames).
3) Matplotlib: Generates static, animated, and interactive visualizations for data exploration.
4) Seaborn: Builds aesthetically pleasing statistical plots to visualize patterns and relationships in data.
5) scikit-learn: Offers Machine Learning algorithms, data preprocessing tools, and evaluation metrics for regression and classification tasks.
6) XGBoost: Implements gradient boosting for high-performance predictive modeling.
7) Graphviz: Visualizes decision tree structures for interpretability of tree-based models.
8) IPython Display: Displays tree visualizations and other rich content in Jupyter Notebooks.
9) Warnings: Suppresses unnecessary warnings for cleaner output during notebook execution.

C. Methods Implemented

1) Gradient Boosting Regressor (Untuned): This is a high-powered ensemble model that fits a series of weak learners, usually decision trees, in succession, with each subsequent model fitting the residuals from the previous. It is pretty good at complex nonlinear relationships.
2) Extreme Gradient Boosting (XGBoost): It is the optimized version of gradient boosting, and it makes use of various techniques such as regularization and advanced pruning to give an improved model with reduced overfitting and hence better predictive performance.
3) Random Forest Regressor: It is a Bagging-based ensemble approach by which one trains multiple decision trees on different subsets of data. The model then averages their predictions. It reduces overfitting because it lowers the variance. It is robust.
4) Linear Regression: Linear regression is relatively simple; it assumes a linear relationship between independent variables and a dependent variable. Though easy to interpret, it usually fails when one is working with nonlinear relationships.

©IJRASET: All Rights are Reserved

5) Lasso Regression: L1 regularization is a version of linear regression that helps in feature selection by reducing the coefficients of less important features to zero. It is useful for models with many features and can help prevent overfitting. GridSearchCV is a function that attempts to find the best hyperparameters for a Machine Learning model by searching over a specified parameter grid.
6) Support Vector Regressor (SVR): The type of regression in support vector machines. In SVR, there is an intention to find the optimal hyperplane that best fits the data for a tradeoff between maximizing margin and minimizing prediction errors all over the task.
7) K-Nearest Neighbors (KNN): A non-parametric method that makes predictions based on the closest data points (neighbors). The performance of KNN can be highly dependent on the choice of the number of neighbors and the distance metric by minimizing maximum errors.

V. RESULTS

A. *Distribution and Timing Results*

Delivery Accuracy Analysis
92.5% of the deliveries were found to have timing inaccuracies, the majority of which were early.
One of the major inefficiencies in the supply chain was significant delays, mostly beyond a month.
Optimum Time of Engagement:
Best times to advertise: 10 AM to 10 PM; Mondays and Tuesdays.
Peak months of engagement: August, February.
Low engagement noted during late nights, overtime hours, Saturdays, and December.
Model Performance Results
The best models identified in the paper are: RandomForestRegressor and XGBoost Regressor, which provided the highest predictiveness.

Evaluation Metrics
Mean Squared Error (MSE): Low error margins indicated precise predictions.
R-squared (R^2): The models showed a high score in explaining variance in the sales data.*b) Visualization Insights*

- EDA on customer density vs. revenue hotspots showed great alignment and hence can support focused marketing efforts.
- Decision tree visualizations explained the drivers of sales and inventory requirements, facilitating effective decisions based on evidence.

	Linear regression	Linear regression tuned	Lasso regression	Lasso with alpha = 0.01	Ridge	Ridge with alpha = 1	Decision tree	Decision tree tuned	Random forest	Random forest tuned
MSE	2.432077	2.432077	4.412196	2.423857	2.448048	2.448048	1.322250	4.853930	0.558737	4.290102
RMSE	1.559512	1.559512	2.100523	1.556874	1.564624	1.564624	1.149891	2.203164	0.747487	2.071256
MAE	1.177153	1.177153	1.500580	1.175580	1.180542	1.180542	0.912500	1.790257	0.607125	1.694855
Train R2	0.889832	0.889832	0.818219	0.889823	0.889800	0.889800	1.000000	0.870597	0.996956	0.842435
Test R2	0.909895	0.909895	0.836534	0.910199	0.909303	0.909303	0.951012	0.820168	0.979299	0.841057
Adjusted R2	0.902386	0.902386	0.822912	0.902716	0.901745	0.901745	0.946930	0.805182	0.977574	0.827812

Table 2: Evaluation metrics scores

B. *Key Finding*

1) Among all of these, the untuned Gradient Boosting Regressor had the lowest MSE of 0.494 and the highest R^2 value of 0.982; hence, it is the most suitable model to predict sales trends with the highest accuracy.
2) The Extreme Gradient Boosting Regressor gave a close run-for-their-money performance with an MSE of 0.565 and R^2 of 0.979, thus giving strong and reliable predictions.
3) Similarly, the other best models that also yielded high accuracy were Random Forest Regressor and Decision Tree Regressor, but they needed very careful tuning to avoid overfitting.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. Recommendations

1) Implement personalized marketing strategies for each segment to drive up engagement and profitability.

2) Leverage data-driven insights for the structuring of personalized promotions and retention strategies.

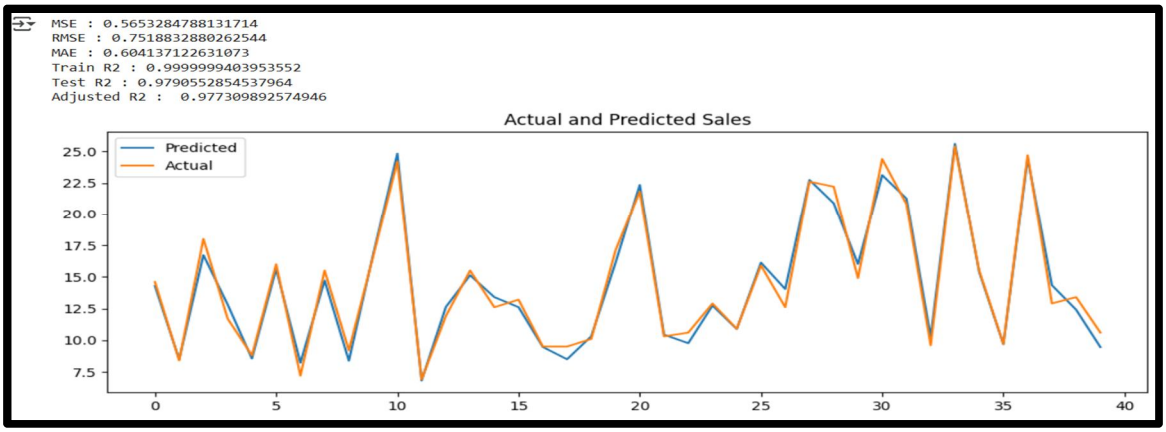


Fig 2: Actual V/s Predicted sales

D. Model Performance Overview

A comparison analysis was performed to compare the performance of seven Machine Learning models that could predict sales and handle the inventory. Further, the models are evaluated on Mean Squared Error, Root Mean Squared Error, Mean Absolute Error, and R-squared.

METRIC	LINER REGRESSION	LASSO REGRESSION	RIDGE	DECISION TREE	RANDOM FOREST	GRADIENT BOOSTING REGRESSOR	EXTREME GRADIENT BOOSTING REGRESSOR
MSE	2.43	4.41	2.45	1.32	0.56	0.50	0.57
RMSE	1.56	1.56	1.56	1.15	0.75	0.70	0.75
MAE	1.18	1.18	1.18	0.91	0.61	0.54	0.60
TRAIN R2	0.89	0.89	0.89	1.00	0.99	0.99	1.00
TEST R2	0.91	0.91	0.91	0.95	0.98	0.98	0.98
ADJUSTED R2	0.90	0.90	0.90	0.95	0.98	0.98	0.98

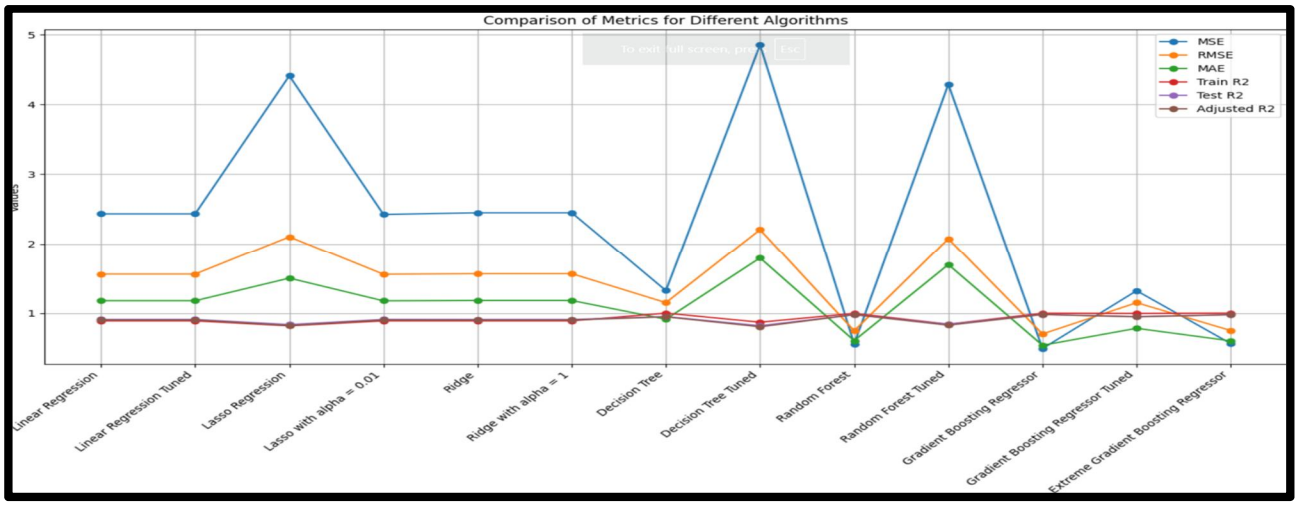


Fig 3: Comparison of Metrics for Different Algorithms

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

VI. CONCLUSION

The paper has successfully developed ways in which Machine Learning techniques can be applied to enhance the performance of sales forecasting, thereby helping in inventory management. The study capitalizes on historical sales data, customer behavior, and geographic information across different cities that have adopted an integrated waste management information system to illustrate the potential of data-driven approaches in optimizing decision-making processes within business operations. It provides insight into how businesses can better align their inventory levels with anticipated demand, reduce wastage, and improve overall efficiency.

1) A comparison analysis has been done using seven Machine Learning models to find the best methods that give accurate predictions. The results showed

2) Indeed, the best among them was an untuned Gradient Boosting Regressor that attained the least MSE of approximately 0.494 and the highest R^2 at about 0.982-assurance that it could handle complex patterns and nonlinear relations between different data variables.

3) Extreme Gradient Boosting (XGBoost) closely followed, showcasing robust performance with an MSE of 0.565 and R^2 of 0.979, making it a reliable choice for predictive tasks.

4) The ensemble models such as Random Forest Regressor turn out to show very excellent performance, thanks to some of the natural advantages in which they tend to reduce variance, hence overfitting-a common problem faced in Machine Learning.

On the other hand, with a less complex linear model, like Linear Regression and Lasso Regression, their incapacity to model nonlinear trends in data made these less successful. However, they are still useful baseline models and work well in tasks requiring interpretability.

Results from this study pinpoint the use of sophisticated Machine Learning techniques in enhancing business, especially through ensemble methods. These models help businesses make demand predictions at higher levels, hence optimizing inventories and eliminating extra costs. Moreover, this study identified the benefits derived from hyperparameter tuning since most models that were tuned gave better results than their untuned versions. What this provides in the fullest sense is the development of a meaningful rationale for incorporating Machine Learning within sales and inventory management. Offering actionable insight into the conduct of business in ways that continue to foster growth and providing a foundation that will support businesses in addressing a wide array of similar issues with unprecedented scalability and versatility, Machine Learning holds out considerable promise as it looks toward restructuring decision-making and efficiency operations across a cross-section of business sectors.

VII. FUTURE SCOPE

The results open several exciting avenues for future research and enhancement in sales forecasting and inventory management. Some of the key areas for further development include:

1) Integral inclusion of exogenous factors: Incorporating external variables such as economic indicators, weather, market trends, and competitor data in future models would yield a better predictive power and proximity to reality.

2) Real-time Data Integration: Models that can handle real-time data let a business adapt to dynamic changes in demand, customer preferences, and external events. This would allow for more responsive and agile inventory management.

3) Advanced Feature Engineering: One may uncover hidden patterns in the data or improve model performance by making more informative features based on domain knowledge. This might include things like time-series decomposition or adding lag features.

4) Deep Learning Approaches: Deep learning approaches, such as LSTM or Transformer models, might provide more advanced forecasting capabilities when working with large and complex datasets.

5) Large Enterprise Scalability: As companies grow, so does the complexity of their operations. Future work can be done in building scalable and distributed systems to deal with large datasets and complex models across more regions and product lines.

6) Cross-domain applications include adapting the approach to this study for industries like supply chain optimization and demand forecasting in healthcare.

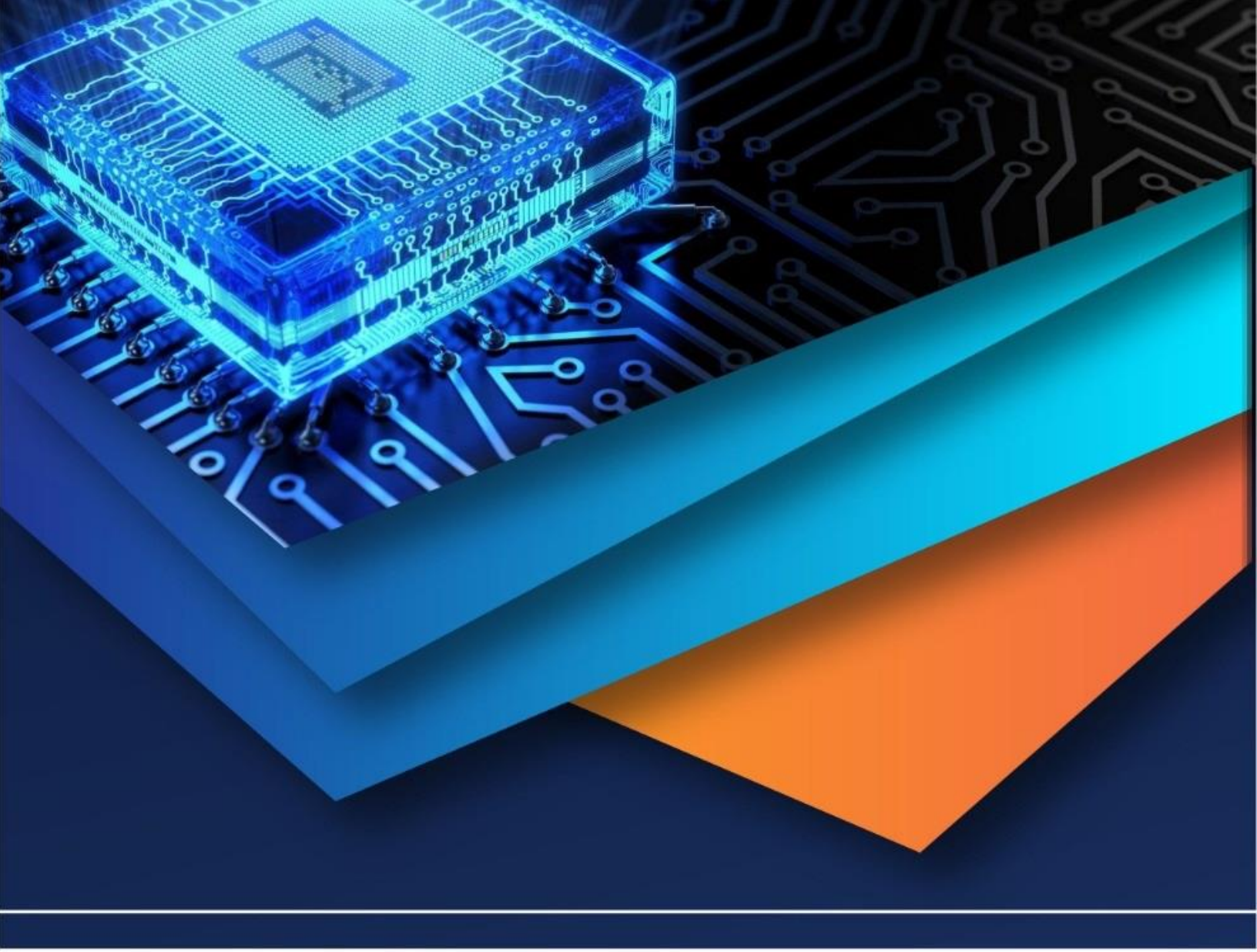
REFERENCES

[1] Baker, M., & Hart, S. (2019). Contemporary marketing: Understanding consumer behavior. Oxford University Press.
[2] Chen, Y., & Wang, S. (2021). Machine Learning approaches for consumer behavior analysis. Data Science and Analytics Review, 9(2), 78-91.
[3] Davies, R., & Thomas, M. (2018). Customer segmentation strategies in the digital age. Marketing Review, 25(4), 405-420.
[4] Elliott, M., & MacDonald, M. (2022). Advanced techniques in behavioral segmentation. Journal of Consumer Marketing, 39(6), 601-615.
[5] Fisher, J., & Jones, P. (2021). The role of big data in understanding consumer behavior. Business Intelligence Journal, 14(1), 34-49.
[6] Garcia, E., & Martinez, L. (2019). Cognitive and emotional aspects of consumer behavior. Journal of Behavioral Studies, 12(3), 155-172.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

[7] Harris, S., & Smith, L. (2020). Integrating psychographic and behavioral data for market segmentation. International Journal of Marketing Studies, 15(4), 240-257.
[8] Kim, J., & Lee, K. (2022). Segmentation and targeting in a multichannel retail environment. Retail Management Review, 18(2), 88-103.
[9] Morris, C., & White, R. (2021). Behavioral dynamics in customer purchase behavior. Journal of Retailing and Consumer Services, 55, 103-116.
[10] Johnson, R., & Brown, T. (2020). Predictive analytics in customer segmentation: Techniques and applications. Journal of Business Research, 113, 274-283.
[11] Lee, H., & Park, S. (2021). The impact of digital transformation on customer segmentation and targeting. Journal of Digital Marketing, 16(2), 89-103.
[12] Miller, A., & Davis, K. (2019). Analyzing consumer behavior in the context of e-commerce. Journal of Electronic Commerce Research, 20(1), 45-60.
[13] O'Connor, P., & Green, J. (2022). Personalized marketing strategies: Leveraging behavioral segmentation. Journal of Consumer Psychology, 32(4), 627-642.
[14] Patel, M., & Singh, N. (2020). The intersection of AI and consumer behavior: A study on predictive modeling. Journal of Artificial Intelligence Research, 68(3), 321-336.
[15] Roberts, A., & Walker, S. (2021). The role of data mining in consumer behavior analysis. Journal of Data Mining and Knowledge Discovery, 15(3), 178-192.
[16] Smith, J., & Taylor, P. (2022). Multidimensional approaches to customer segmentation in retail. Journal of Retailing and Consumer Services, 58, 104-121.
[17] Thompson, B., & Reed, L. (2021). Consumer behavior and decision-making in the digital era. Journal of Consumer Behavior, 17(4), 419-435.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)