DOI: https://doi.org/10.3126/ljbe.v12i2.77417

Vol. XII, No. 2 Journal homepage: www.nepjol.info/index.php/ljbe April, 2025

The Lumbini Journal of

Business and Economics

Peer Reviewed

Trending Sales in E-commerce Using Machine Learning: A Study Referring to Mobile Phone Set

Biswajit Biswas[1]
Sourav Das[2]

Article Info

Received:
27 December 2024

Revised:
15 February 2025

Accepted:
18 February 2025

Abstract

Purpose:*This is an investigation to fund the influence of price, technical features,and marketing strategies, on the sales volume of mobile phone sets and developed a model for mobile phone companies using machine learning techniques.*

Method: *Starting from simple linear regression, this research work attempted to use decision trees and random forests for its investigation as a part of machine intelligence tools. The models are compared based on performance metrics such as mean squared error, R-squared, and precision-recall curves.*

Results: *The results indicate that machine learning models, particularly the random forest algorithm, can accurately forecast mobile phone sales with a good accuracy rate.*

Conclusion: *This research work contributes to the existing literature on machine learning in general and sales forecasting in particular. It is providing insights and recommendations for future research exploring the other techniques under machine intelligence for the predictive modelling. The findings have practical implications for mobile phone companies seeking to optimize their production, marketing, and pricing strategies based on accurate sales predictions.*

Keyword: *S*ales prediction, machine learning, mobile phone.

I. Introduction

Sales prediction is a vital part of business planning for mobile phone companies. Accurate sales forecasting helps companies optimize their production, marketing, and pricing strategies to maximize profits and minimize costs. Traditional sales forecasting methods are often limited by their inability to handle large and complex datasets. In the present age, machine learning (ML) techniques have emerged as a promising approach for sales prediction due to their ability to learn patterns and relationships with data and make truthful predictions. The objectives of thiswork are to develop a sales prediction model for mobile phone companies using machine learning (ML) techniques. The work will use historical sales data and other relevant variables like price, features, campaigns, and consumer preferences to identify the most significant

[1] *Biswajit Biswas, Ph.D. is an Assistant Professor at Department of Business Administration University of Kalyani, India. Email: bbmbaku17@klyuniv.ac.in*

[2] *Saurav Das is a Marketing Manager at Punjab National Bank Circle Office, Kolkata – 700124 Barasat, India. comtia.sourav@gmail.com*

predictors of mobile phone sales. Various machine learning (ML) algorithms, including linear regression, decision trees, and random forests, will be applied to the dataset to compare and develop predictive models. The performance of each model will be evaluated based on several metrics, like mean squared error, R-squared, and precision-recall curves. The findings will provide meaningful insights for practitioners and researchers seeking to improve their sales prediction capabilities using machine learning (ML) techniques. Additionally, this work has practical implications for mobile phone companies seeking to optimize their sales forecasting accuracy and develop effective business strategies.

Objective of the work

- To develop and compare machine learning algorithms for predicting sales of mobile phones.
- To identify the most accurate algorithm for this task by analysing historical sales data, and to explore the potential of ML in improving sales forecasting in the mobile phone industry.
- To provide valuable insights to mobile phone companies seeking to optimize their production, marketing, and pricing strategies

II. Review

The trend of E-commerce sales attracted customers and sales forecasting plays a significant role in organizational growth. Many authors used statistical analysis for prediction but the present trend suggests that ML for more information and better decision-making. It predicts accurately and gives more accuracy. Sales prediction gives a way for the future and also provides a way for the organization's performance (Xavieret al., 2023). Sales prediction is important for an organization to sustain itself in the competitive market. In this competitive environment, AI helps the organization gain insight into sales data and makes situational decisions. In an e-commerce site, an organization collects the customer's reviews on specific products which help the organization to know the customer's preferences, and needs. In this paper, authors used ANN to develop a sales forecasting model and they used the customer review as input data (Biswas et al., 2023).

The rapid growth of e-commerce has made understanding customer behavior and purchase intent crucial for online retailers. This review explores the factors influencing online shopping and the application of ML techniques to predict buying intentions. Predicting buying intention in online shopping is a complex task that requires a comprehensive understanding of customer behavior and the application of advanced machine learning techniques. By identifying key factors influencing purchase decisions and leveraging data-driven insights, online retailers can improve their marketing strategies, enhance the customer experience, and increase sales (Islam et al., 2023). The increasing popularity of e-commerce has made understanding online shoppers' purchasing intentions a crucial area of research.

Businesses must identify factors influencing purchase decisions to optimize their marketing strategies and enhance sales. This paper proposes an ensemble learning-based model to analyze online shoppers' purchase intentions, aiming to improve prediction accuracy and inform targeted marketing efforts. The proposed model demonstrates superior performance compared to existing methods, highlighting the benefits of combining multiple models. By understanding the factors influencing purchase decisions, businesses can tailor their marketing strategies to increase conversion rates and enhance customer satisfaction (Karakayaet al., 2023). The e-commerce landscape has witnessed rapid growth, driven by technological advancements. Data analytics and machine learning (ML) have emerged as crucial tools for businesses to optimize operations, enhance customer experiences, and gain a competitive edge. This review examines the impact of these technologies on e-commerce, focusing on their applications, benefits, challenges, and the specific context of the retail

sector. Data analytics is pivotal in understanding customer behavior, market trends, and optimizing business processes. Machine learning (ML) algorithms enable computers to learn from data and improve their performance over time (Rajan et al., 2022). Ranjansaid that this article analyses the market price of different mobile brand, assessing factors such as wear resistance, charging time, design, memory size, battery life, communication stability, drop resistance, camera quality, and the likelihood of repurchasing as dependent variables. Many researchers have employed similar methods to investigate the current state and future trends of various products, yielding a range of findings utilizing mathematical statistics and linear regression algorithms.

For this paper, author used support vector machines to forecast the sales predictions of different mobile phones brand, with sales forecast levels serving as the output variables (Duan et al., 2019).Cheriyansaid that estimating the price of a mobile device, many factors need to be considered. A recent study found that when dealing with extensive data sets, the SVM technique is superior to others, such as multiple linear regressions, for price prediction accuracy. The research also concluded with comparative results from both feature selection algorithms and classifiers, except for the Wrapper attribute Eval and Decision Tree J48 classifier combination. This combination yielded maximum accuracy while selecting only the most appropriate features. It's worth noting that adding irrelevant or redundant features to the data set during Forward selection can decrease the efficiency of both classifiers (Cheriyan et al., 2018). Bajaj said that the goal of this work is to predict sales patterns and product quantities using key features extracted from raw data. Through analysis and exploration of the collected data, Authors aim to gain a comprehensive understanding of the information. In order to enhance revenue growth for businesses, traditional methods have proven to be insufficient, prompting the utilization of Machine Learning techniques to shape business strategies based on consumer purchasing behaviour (Bajaj et al., 2020).

There has been significant interest among practitioners in using ML techniques to predict customer retention and profitability. These methods involve analyzing a customer's past purchase behavior to extract hidden characteristics, which are believed to be the result of an underlying stochastic process. This approach, known as the characteristics approach, identifies customer features such as the frequency of past transactions; the time elapsed since the last transaction, and the change in total spending. The study assesses the performance of constructed estimators by using classifiers like logistic Lasso, extreme learning machines, and gradient tree boosting. The study also introduces a framework for dynamically predicting a customer's purchase behavior in the next month in a non-contractual setting, employing data-driven methods and various machine learning algorithms. The analysis considers a month as the time unit, with purchase values being summed in cases where a customer has multiple purchases in a month (Martinez et al., 2020).

In the fastest-growing e-commerce market review analysis takes an important role in online sales, customer retention, potential buyer, etc. In this paper, the authors considered the three largest e-commerce sites for review analysis. The authors used the feed-forward ANN supervised learning to develop the model (Biswas et al., 2021). Soumakassaid that the need for computer s to support sales prediction is highlighted in this paper, as it can aid decision-making and provide estimates of future sales. The focus of this literature review is on the use of ML techniques for predicting sales food, an vital task for various businesses, including restaurants, supermarkets, bakeries, and patisseries. The second section discusses the model of sales prediction as a ML task, along with the learning techniques that have been employed in the past.

One of the challenges is that each product may have a separate prediction task, which requires different learning algorithms to match the assets of each time sequences. While human decision-makers are allowed to enter external variables to adjust for unknown factors that could affect demand, there is potential for improvement by incorporating sales information for similar products using multi-target regression techniques. The paper suggests that this area has not been fully explored in the past, and could lead to greater accuracy by exploiting

dependencies among similar products (Tsoumakas, 2019).Computingtechnology increases its application in the business day by day for its accuracy in business decision-making. Soft intelligence approaches minimize the decision-making time and decide on a real-time basis depends on online real-time data.

To survive in this technology-based competitive age every organization spreads their business wings in e-commerce platforms (Biswas & Sanyal, 2019). In this research article, three machine learning (ML) algorithms - Random Forest, Gradient Boosting, and KNN - were utilized to predict sales. The results were assessed using the mean absolute error (MAE) metric. The Random Forest algorithm demonstrated the best performance among the three, achieving a lower MAE of 0.409178. It was observed that the training and cross-validation scores for all three models improved as the training size increased. The unexpected poor performance of the Gradient Boosting model was noted, especially considering its effectiveness in previous sales forecasting studies. The study also highlighted that the small dataset size may have contributed to the high errors observed in the models. Furthermore, it identified the most important features used by the Random Forest and Gradient Boosting models and suggested that acquiring more data could enhance the predictive power of the models. In conclusion, the study suggests that implementing machine learning algorithms for sales forecasting holds promise for significant improvements in revenue and customer service (Odegua, 2020).

Various supervised learning algorithms and outlines the basic steps involved in the machine learning process. A common type of machine learning task is the classification problem, where the goal is to learn a function that maps an input vector to one of several pre-defined classes based on labelled input-output examples .Supervised algorithms rely on predetermined classes that are defined by humans and applied to labelled segments of data. During the training phase, the learning algorithm processes input samples from the training data and learns the underlying features to construct the learning model. Supervised techniques have shown to be more effective than unsupervised techniques due to the availability of considered training data, which offers clear optimization conditions for model development (Nasteski, 2017).

In this work author predict the prediction model was trained using the Mobile Phone Price Class dataset from the Kaggle data science community website which classifies mobiles into Mobile phone brand. Python is chosen because its ML libraries are easily accessible. The model was trained using a variety of classification algorithms in an effort to identify the one that can more reliably predict the pricing class for mobile devices

III. Methodology

The author conducted a statistical analysis using parameters such as Brand, Model, Color, Memory, Storage, Rating, Original Price, and Selling Price. This analysis provides insight into the impact of reviews on online shopping. Author used secondary data for this work. The analysis was carried out in Jupyter Notebook using machine learning techniquesfollowed by the section VI data analysis and interpretation.

Data Analysis and Interpretation

The Python packages used in the model are imported first for data analysis and prediction. The numpy package is used for mathematical operations, pandas for data analysis, matplotlib. pyplot package for plotting histograms and scatter plots, and seaborn package is used for visualizing the data.

Code for package import is written as

```
Import numpy as pd

Import pandas as pd
```

Import matplotlib.pyplot as plt

Import seaborn as sns

From the above code the dataset is imported, and the shape of the dataset is checked to ensure that the data is imported properly for analysis.

dataset = pd.read_csv('Flipkart_Mobiles (2).csv')

dataset.shape

dataset.head(5)

Table 1

Import of Dataset

Brand	Model	Color	Memory	Storage	Rating	Selling Price	Original Price
OPPO	A53	Moonlight Black	4 GB	64 GB	4.5	11990	15990
OPPO	A53	Mint Cream	4 GB	64 GB	4.5	11990	15990
OPPO	A53	Moonlight Black	6 GB	128 GB	4.3	13990	17990
OPPO	A53	Mint Cream	6 GB	128 GB	4.3	13990	17990
OPPO	A53	Electric Black	4 GB	64 GB	4.5	11990	15990

In Table 1 author has used isnull.() command to detect the null value from the entire dataset, when the null values get detected author drop out the null value from the given set of data by using dropna.() command and display the remaining data below.

Table 2

Drop Null Value and Create New Dataset

	Rating	Selling Price	Original Price
count	2897	2897	2897
mean	4.2413	24692.49	26656.68
std	0.2728	26373.59	28251.23
min	2.3	1000	1000
25%	4.1	9990	10389
50%	4.3	14999	16300
75%	4.4	27499	29999
max	5	179900	189999

In table 2, the data is presented as a statistical summary of the data frame. If the dataset contains numerical values, the `describe () ` command is used to provide a graphical representation that measures the descriptive statistics of the data.

Table 3

Adding New Column

Brand	Model	Color	Memory	Storage	Rating	Selling Price	Original Price	Mobile	Discount	Discount in %
OPPO	A53	Moonlight Black	4 GB	64 GB	4.5	11990	15990	OPPO A53	4000	25.01564
OPPO	A53	Mint Cream	4 GB	64 GB	4.5	11990	15990	OPPO A53	4000	25.01564
OPPO	A53	Moonlight Black	6 GB	128 GB	4.3	13990	17990	OPPO A53	4000	22.23458
OPPO	A53	Mint Cream	6 GB	128 GB	4.3	13990	17990	OPPO A53	4000	22.23458
OPPO	A53	Electric Black	4 GB	64 GB	4.5	11990	15990	OPPO A53	4000	25.01564

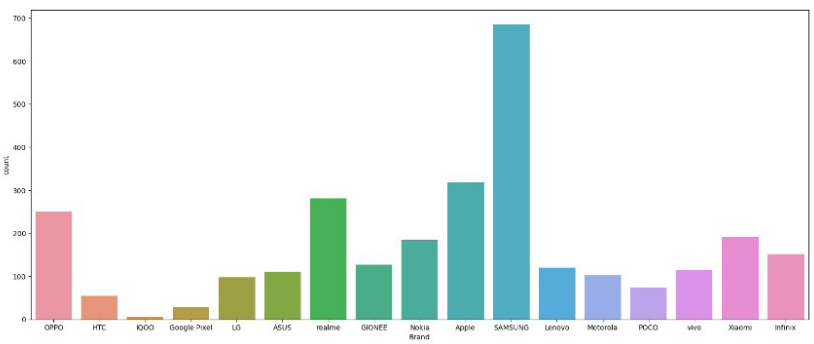
Code used:

```
fc5 = new_dataset.copy()
fc5['Discount in %'] = (fc5['Discount'] / fc5['Original Price']) * 100
fc5
```

In Table 3 after completing described and displayed a statistical analysis of the data frames. Author added Discount and Discount in percentage column for further visualisation of the mobile phone sales prediction.

Figure 1

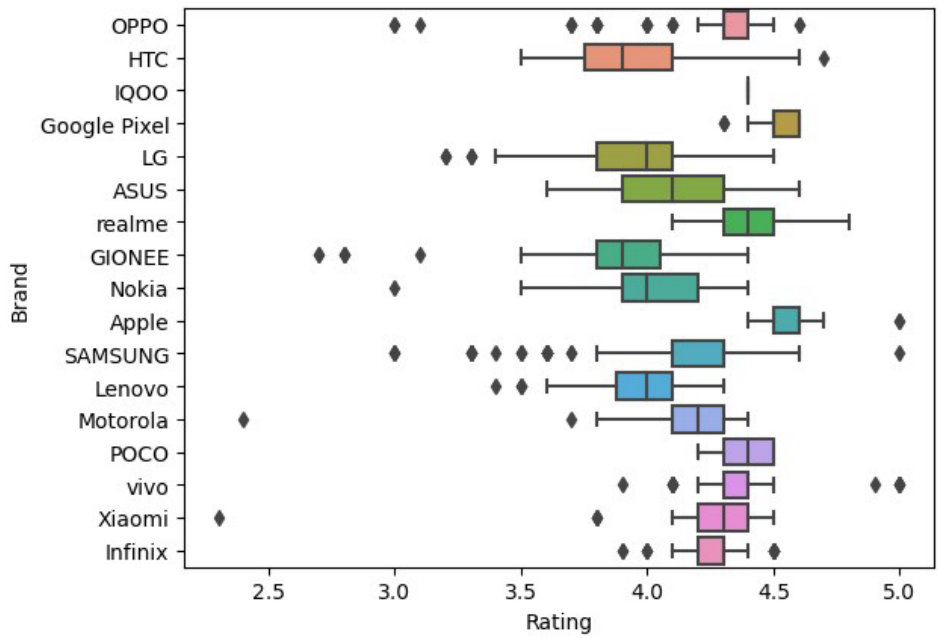
Frequency of Different Brand



In Figure 1 author expressing the frequency level of selling of mobile phones from different brands. Where among all the mobile brands Samsung has the higher frequency level than other brands.Which means the mobile phones from Samsung brand were sold more than others.

Figure 2

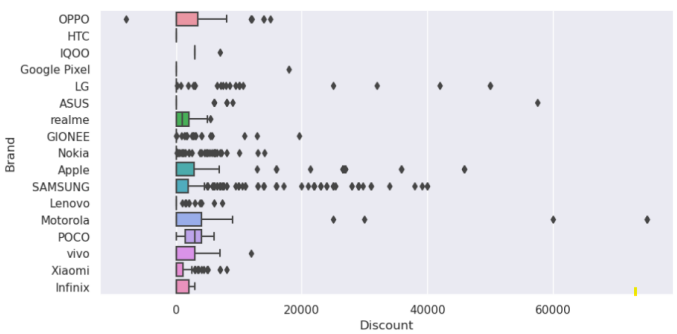
Brand VS Rating



In Figure 2 author displaying the customer ratings of different brands. Where the rating of Google is 3.4 and one plus has the rating within 3.9-4.0, Apple has a rating between 3.1-4.5, Huawei has a rating between 3.1-4.1, Samsung has a rating between 3.3-3.5, Sony has a rating between 3.1-3.9, Motorola has a rating between 3.0-3.4 and Nokia has 3.1-3.4 rating. Which indicates that Motorola has the lowest ratings among all the brands, where Huawei has the highest ratings than others.

Figure 3

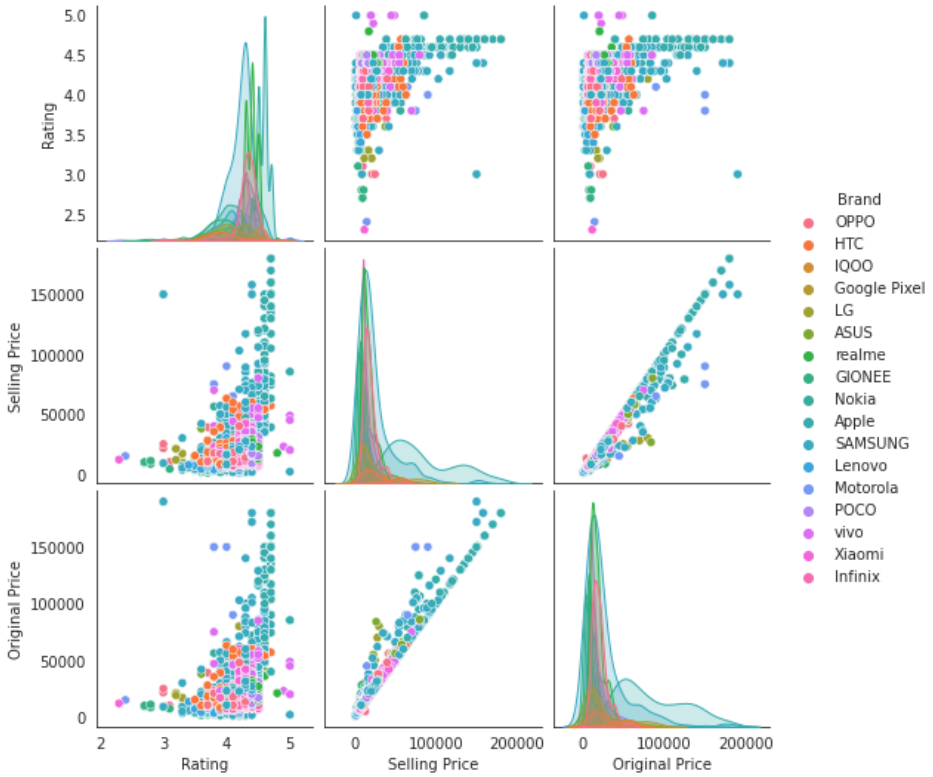
Brand VS Discount



In Figure 3 author displaying the customer discount of different brands. Here author can see that the most discount provider mobile phone brand is Motorola and author also see that the lowest discount provider mobile phone brand is HTC, IQOO and Lenovo.

Figure 4

Pair Plot



In Figure 4 author displaying the pair plot for visualize pairwise comparison each and every variables. Here author considered three variables Rating, Selling Price and Original Price each variables compare with other two variables and after comparison author gets this pair plot for better understanding.

In Figure 5 and Figure 6 author displaying the customer original price of different brands, also selling price of different brand. Author displaying the both diagram for understanding the different between original price and selling price. After undestanding the difference between bothdiagram author can easily perdic the amount of discount of different brands.

Figure 5

Original Price Count

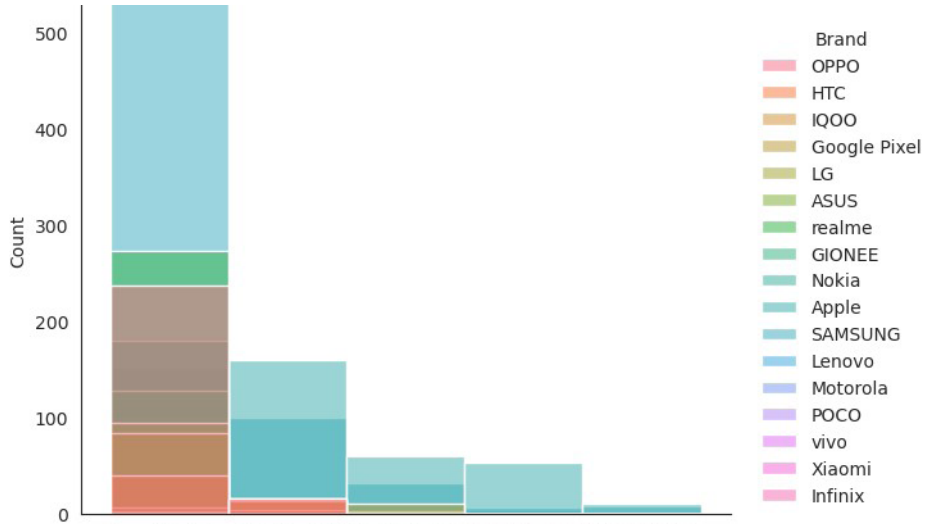
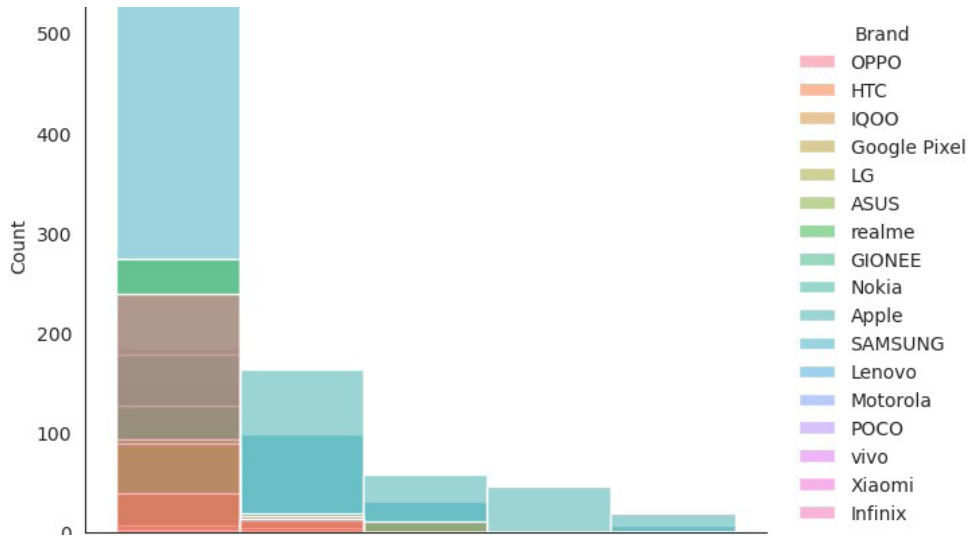


Figure 6

Selling Price Count



IV. Conclusion and Implication

The work focused on using machine learning to predict mobile phone sales. It showed that machine learning algorithms can effectively predict sales, with the Random Forest algorithm being the most accurate. The work emphasized the importance of data quality and quantity, as well as considering external factors such as consumer preferences and economic conditions. However, the study has some limitations that need to be addressed in

future research, including exploring additional features and data sources, using deep learning techniques, and including a wider range of mobile phone brands and models. Overall, the work has significant implications for mobile phone companies aiming to improve their sales forecasting and optimize their production, marketing, and pricing strategies. By incorporating machine learning techniques into their decision-making processes, companies can gain a competitive advantage in a dynamic and competitive market.

The work focuses on predicting mobile phone sales using machine learning. It has significant managerial implications for mobile phone companies aiming to enhance their production, marketing, and pricing strategies. Overall, the model offers a meaningful insight into the application of machine learning (ML) algorithms for sales forecasting and underscores the potential of this technology in improving business decision-making within the mobile phone industry.

Reference

Bajaj, P., Ray, R., Shedge, S., Vidhate, S., & Shardoor, N. (2020).Sales prediction using machine learning algorithms. *International Research Journal of Engineering and Technology (IRJET)*, *7*(6), 3619-3625.

Biswas, B., Sanyal, M. K., & Mukherjee, T. (2021). Feedback analysis for digital marketing in India: Empirical study on Amazon. in, Flipkart, and Snapdeal. *International Journal of Online Marketing (IJOM)*, *11*(1), 78-88.

Biswas, B., &Sanyal, M. K. (2019, January). Soft intelligence approaches for selecting products in online market. In *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 432-437).IEEE.

Biswas, B., Sanyal, M. K., & Mukherjee, T. (2023).AI-based sales forecasting model for digital marketing. *International Journal of E-Business Research (IJEBR)*, *19*(1), 1-14.

Cheriyan, S., Ibrahim, S., Mohanan, S., & Treesa, S. (2018, August).Intelligent sales prediction using machine learning techniques.In *2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE)* (pp. 53-58).IEEE.

Duan, Z., Liu, Y., & Huang, K. (2019, May). Mobile phone sales forecast based on support vector machine. In *Journal of Physics: Conference Series* (Vol. 1229, No. 1, p. 012061). IOP Publishing.

Islam, M. S., Naeem, J., Emon, A. S., Baten, A., Al Mamun, M. A., Waliullah, G. M., ...& Mridha, M. F. (2023, June). Prediction of buying intention: factors affecting online shopping. In *2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM)* (pp. 1-6). IEEE.

Karakaya, A., Karakaya, İ., & Temizceri, T. (2023, December). An online shoppers purchasing intention model based on ensemble learning. In *2023 4th International Informatics and Software Engineering Conference (IISEC)* (pp. 1-4). IEEE.

Martinez, A., Schmuck, C., PereverzyevJr, S., Pirker, C., &Haltmeier, M. (2020).A machine learning framework for customer purchase prediction in the non-contractual setting. European Journal of Operational Research, 281(3), 588-596.

Mboungou, M. M. B., Yamin, I., Zhang, S., & Iqbal, A. (2023, November). Sentiment analysis of client reviews on a French e-commerce platform. In *2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE)* (pp. 1-6).IEEE.

Nasteski, V. (2017).An overview of the supervised machine learning methods.Horizons.b, 4, 51-62.

Odegua, R. (2020). Applied machine learning for supermarket sales prediction. project: Predictive machine learning in industry.

Rajan, S. D., Vavilapalli, S., Hasan, S., Kumar, R., Rafa, N., & Muda, I. (2022, December). A survey on the impact of data analytics and machine learning techniques in e-commerce. In *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)* (pp. 1117-1122). IEEE.

Tsoumakas, G. (2019). A survey of machine learning techniques for food sales prediction. *Artificial Intelligence Review*, *52*(1), 441-447

Xavier, K. A., Manjunath, C., Manohar, M., Gurudas, V. R., Jayapandian, N., &Balamurugan, M. (2023, April). Analytical Methods of Machine Learning Model for E-Commerce Sales Analysis and Prediction.In *2023 IEEE International Conference on Contemporary Computing and Communications (InC4)* (Vol. 1, pp. 1-6).